



# Towards Efficient and Scalable Data-Intensive Content Delivery: State-of-the-Art, Issues and Challenges

Irene Kilanioti<sup>1</sup>(✉), Alejandro Fernández-Montes<sup>2</sup>, Damián Fernández-Cerero<sup>2</sup>,  
Anthony Karageorgos<sup>4</sup>, Christos Mettouris<sup>1</sup>, Valentina Nejkcovic<sup>3</sup>,  
Nikolas Albanis<sup>4</sup>, Rabih Bashroush<sup>5</sup>, and George A. Papadopoulos<sup>1</sup>

<sup>1</sup> Department of Computer Science, University of Cyprus, Nicosia, Cyprus  
irenekilanioti@gmail.com, {mettour, george}@cs.ucy.ac.cy

<sup>2</sup> Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla,  
Seville, Spain  
{afdez, damiancerero}@us.es

<sup>3</sup> Faculty of Electronic Engineering, University of Nis, Niš, Serbia  
valentina@elfak.ni.ac.rs

<sup>4</sup> TEI of Thessaly, Karditsa, Greece

karageorgos@computer.org, nikosbit@gmail.com

<sup>5</sup> University of East London, London, UK  
r.bashroush@qub.ac.uk

**Abstract.** This chapter presents the authors' work for the Case Study entitled "Delivering Social Media with Scalability" within the framework of High-Performance Modelling and Simulation for Big Data Applications (cHiPSet) COST Action 1406. We identify some core research areas and give an outline of the publications we came up within the framework of the aforementioned action. The ease of user content generation within social media platforms, e.g. check-in information, multimedia data, etc., along with the proliferation of Global Positioning System (GPS)-enabled, always-connected capture devices lead to data streams of unprecedented amount and a radical change in information sharing. Social data streams raise a variety of practical challenges: derivation of real-time meaningful insights from effectively gathered social information, a paradigm shift for content distribution with the leverage of contextual data associated with user preferences, geographical characteristics and devices in general, etc. In this article we present the methodology we followed, the results of our work and the outline of a comprehensive survey, that depicts the state-of-the-art situation and organizes challenges concerning social media streams and the infrastructure of the data centers supporting the efficient access to data streams in terms of content distribution, data diffusion, data replication, energy efficiency and network infrastructure. The challenges of enabling better provisioning of social media data have been identified and they were based on the context of users accessing these resources. The existing literature has been systematized and the main research points and industrial efforts in the area were identified and analyzed. In our works, in the framework of the Action, we came up with

potential solutions addressing the problems of the area and described how these fit in the general ecosystem.

## 1 Introduction

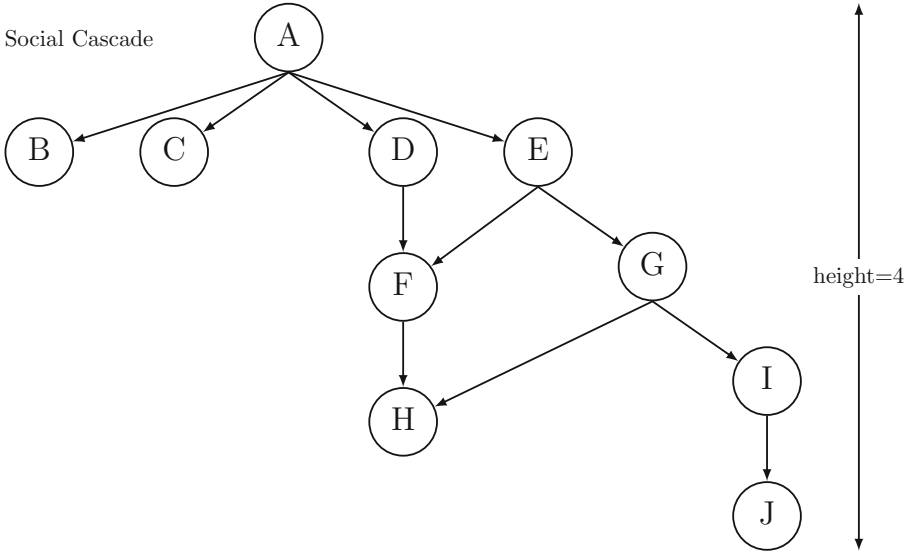
### 1.1 Social Data Streams Features

Herein some basic terminology for the topic of our Case Study entitled “Delivering Social Media with Scalability” within the framework of High-Performance Modelling and Simulation for Big Data Applications (cHiPSet) COST Action 1406 is introduced. The terminology appears in published works [111] and [108], as well.

Social networks, media and platforms enable communication, exchange, business and knowledge acquisition as well as social network users connection with each other with the purpose of sharing content. Social data is the information that social media users share, e.g. check-in information, multimedia data, tags, annotations, and likes, and may include metadata such as the user’s location, native language, biographical data and shared links, whereas ‘streams’ denotes various approaches have been performed that we do not refer to static datasets, but rather to dynamic information generated and transmitted over the Online Social Network (OSN).

Formally, an OSN is depicted by a directed graph  $G = (V, E)$ , where  $V$  is the set of the vertices of the graph representing the nodes of the network and  $E$  are the edges between them, denoting various relationships among the edges of the graph [69]. The semantics of these edges vary, and their interpretation is expanded for various OSNs from personal acquaintance, to common interests, microblogging services or business contact. As far as the directionality of the edges of the social graph is concerned, it is associated with the concept of the OSN: for Facebook, an edge denotes mutual friendship between the endpoints of a link, for Twitter, if the edge between A and B points at B, A’s posts (tweets) appear in B’s main Twitter page, and so on. A social node centrality is indicative of the importance of a node within a social network. It is given in terms of a real-valued function on the vertices of a graph, where the values produced are expected to provide a ranking which identifies the most important nodes [40, 41].

In Rogers’ classic work [150], the author defines *information diffusion* as the process in which an innovation is communicated through certain channels over time among the members of a social system. In this context, the innovation is defined as the first spread of information from an originator. A *social cascade* is a specific case of information diffusion and practically occurs within an OSN, when a piece of information is extensively retransmitted after its initial publication from a user. Cascades can be represented as rooted directed trees where the initiator of the cascade is the root of the tree [26] and the length of the cascade is the height of the resulting tree. Each vertex in the cascade tree can have the information of the user, and the identity of the item replicated in the cascade. Figure 1 depicts an example of the evolution of a social cascade in a directed



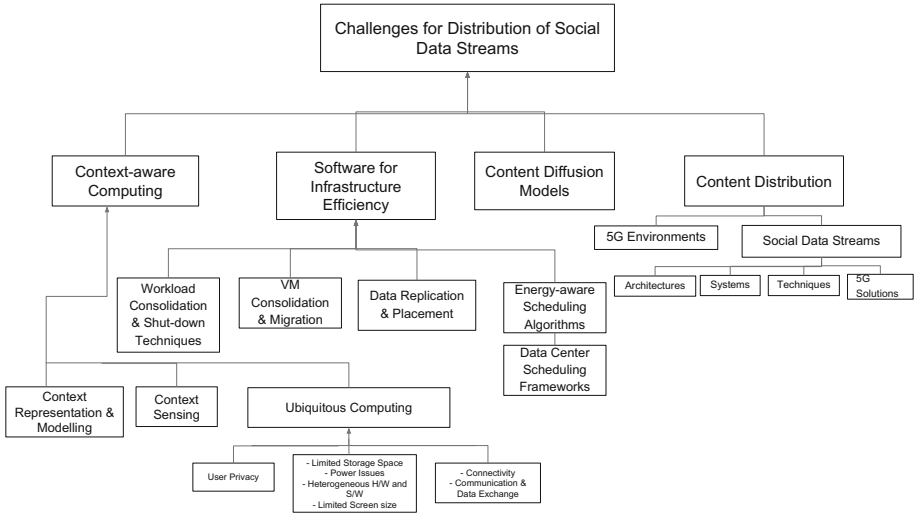
**Fig. 1.** The evolution of a social cascade in Twitter

graph. The cascade follows the arrows' direction. For example, in Twitter, B, C, D, E are followers of A, whereas the adopters of a new information piece could be the nodes, that after having been exposed in a video link, they retransmit it, contributing remarkably to Internet traffic [1].

## 1.2 Challenges for Distribution of Social Data Streams

In the survey [108] we wrote in the framework of (cHiPSet) COST Action 1406 we identified the challenges of enabling better provisioning of social media data based on the context of users accessing these resources. In our works [109–111], that we produced in the framework of the Action, we came up with potential solutions addressing the problems of the area and described how these fit in the general ecosystem.

Distributing social data streams largely depends on the exploitation of usage patterns found in OSNs, and can be improved either through the selective prefetching of content (**cost-effectiveness**) or through the strategic placement/selection of the employed infrastructure (**energy-efficiency**). The cost of scaling such content might be the number of replicas needed for a specific source or it may take into account the optimal use of memory and processing time of a social-aware built system. Optimization of energy efficiency for data centers that support social data interaction and analysis includes tasks such as data growth, data center federation and Content Delivery Network (CDN)-load-balancing at data center level. In our taxonomy (Fig. 2), pillars associated with cost-effectiveness include Context-aware Computing, Content/Information Diffusion Models and Content Distribution challenges, whereas Software for



**Fig. 2.** Taxonomy of challenges for distribution of social data streams.

Infrastructure Efficiency is associated with energy-efficiency. This taxonomy includes solutions or approaches to the ‘Challenges for Distribution of Social Data Streams’. These solutions or approaches require enough effort, hence they can also be considered as a challenge for the research community.

**Context-Aware Computing:** Application of social contextual information, such as profiles, images, videos, biometrical, geolocation data and local data, in situations where conventional bandwidth-intensive content scaling is infeasible, could largely facilitate: the spreading of information, the identification of potential information sources, as well as a paradigm shift in the way users access and control their personal data. User-generated multimedia content is especially difficult due to its long tail nature, with each item probably not popular enough to be replicated in a global scale, but with the long-tail altogether getting sufficient accesses [20]. Social analysis tasks interweaved with context-aware computing could pave the ground for proactive caching mechanisms in the framework of a content delivery infrastructure of streaming providers.

**Software for Infrastructure Efficiency:** The industry has made several efforts to address challenges associated with optimization of energy efficiency for data centers that support social data interaction and analysis [42, 129, 163] such as data growth, isolation, real-time interactions, data center federation and CDN-load-balancing at data center level, but usually lacks from focusing on energy consumption of the employed infrastructures. The challenges in the area of energy-efficient data-centers include workload consolidation and shut-down techniques, Virtual Machines (VMs) consolidation and migration, data replication and placement, and energy-aware scheduling algorithms.

**Content/Information Diffusion Models:** Prevalence of OSNs has transformed the landscape of content exchange. Popularity of relatively data heavy multimedia user generated content (UGC) has also risen [6], resulting in data deluge across all media platforms [12, 127, 141]. Measurement studies, such as [48], attribute the recent increases in HTTP/HTTPS traffic to the extended use of OSNs [28, 49, 69]. Elaborate data manipulation presupposes coping with the size of social graphs with billions of nodes and edges [174]. Facebook, for example, reported that had 1.47 billion daily active users on average for June 2018 and 2.23 billion monthly active users as of June 30, 2018 [4]. Its custom built-in data warehouse and analytics infrastructure [11] has to apply ad-hoc queries and custom MapReduce jobs [55] in a continuous basis on over half a petabyte of new data every 24 h for the creation of meaningful aggregations and analysis. It is also acknowledged that a large proportion of bandwidth-intensive media is distributed via reposted OSN links, contributing significantly to Internet traffic [1], [46]. These challenges are closely associated with the Content/Information Diffusion Models used to represent the diffusion of information over OSNs and facilitate relevant algorithmic solutions (Fig. 2).

**Content Distribution:** The delivery infrastructure of video operators is made up of scattered geo-distributed servers, which with specific cache selection mechanisms direct users to the closest servers hosting the requested data. Transmission Control Protocol (TCP), however, is subject to delay jitter and throughput variations and clients are required to preload a playout buffer before starting the video playback [111]. Thus, the quality of experience (QoE) of media platform users is primarily determined by stalling effects on the application layer. For the YouTube case cache server selection is also highly Internet Service Provider (ISP)-specific, with geographical proximity not being the primary criterion and DNS level redirections for load-balancing purposes occurring quite frequently and substantially contributing to the initial startup delay of the playback. Several network-level and client-level approaches are focused on the detection of such interruptions, that negatively affect the user experience [94]. With the growing popularity of OSNs and the increased traffic due to outspread of information via the latter, the improvement of user experience through scaling bandwidth-demanding content largely depends on the exploitation of usage patterns and geolocation data associated with OSNs. These challenges are closely associated with the Architectures, Systems and Techniques within the 5G infrastructure.

Some key factors contributing to the problem of diffusion of bandwidth-intensive media content over OSNs are discussed below.

**Large-Scale Datasets.** In order to harness the power of social networks diffusion over CDN infrastructure, the key areas of interest that need to be explored include the large size of the graphs, and also the fact that diffusion of links is multiplied through dissemination over sites like YouTube, and amplified by the proliferation of smartphones and cheap broadband connections. The amount of information in OSNs is an obstacle, since elaborate manipulation of the data

may be needed. An open problem is the efficient handling of graphs with billions of nodes and edges.

The desired scaling property refers to the fact that the throughput of the proposed approaches should remain unchanged with the increase in the data input size, such as the large datasets that social graphs comprise and the social cascades phenomena that amplify the situation. Cost of scaling such content can be expressed in different ways. For instance, it may be matched with the number of replicas needed for a specific source. Future experimentations may take into account the optimal use of memory and processing time of an OSN-aware built system.

Internet of Things (IoT) is a global infrastructure that interconnects things based on interoperable information and communication technologies, and through identification, data capture, processing and communication capabilities enables advanced services [5]. Things are objects of the physical world (physical things, such as devices, vehicles, buildings, living or inanimate objects augmented with sensors) or the information world (virtual things), capable of being identified and integrated into communication networks. It is estimated that the number of Internet-connected devices has surpassed the human population in 2010 and that there will be about 50 million devices by 2020 [9]. Thus, the still ongoing significant IoT innovation is expected to generate massive amounts of data from diverse locations, that will need to be collected, indexed, stored, and analyzed.

**OSN Evolution.** Existent works examine valuable insights into the dynamic world by posing queries on an evolving sequence of social graphs (e.g. [146]). Time evolving graphs are increasingly used as a paradigm for the emerging area of OSNs [71]. However, the ability to scalably process queries concerning the information diffusion remains to a great extent unstudied. With the exception of sporadic works on specialized problems, such as that of inference of dynamic networks based on information diffusion data [149], at the time of writing the authors are not aware of relative studies on the information diffusion through OSNs under the prism of graphs dynamicity.

**5G Approaches.** The demand for high-speed data applications that has risen in recent decade lead to development of Fifth Generation Wireless (5G) communications. Development of efficient mechanisms for supporting mobile multimedia and data services is prerequisite for 5G networks. Real bottleneck of today's mobile networks is the radio access network and the backhaul. Caching in the intermediate nodes, servers, gateways, routers, and mobile users' devices can reduce doubled transmission from content providers and core mobile networks.

Known caching techniques that can be used within 5G are: content distribution network, information-centric networks, content-centric networking, http web caching, evolved packet core caching, radio access network caching, device to device caching, proactive caching, predictive caching, cooperative caching [23]. Those techniques are using different algorithms and models. Analysis presented

in [23] has shown that the deployment of those caching techniques in mobile network can reduce redundant traffic in backhaul, minimize the traffic load, increase the transfer rate in mobile network and reduce the latency. Correlation of several caching methods and procedures could result in improving network performance and obtaining better results.

On the other hand, well known bottleneck that 5G brings is the complex heterogeneity of the network. Particularly, network consists of different technologies that coexist, where some technologies could potentially disable the transmission of data of equipment that use other technologies. Thus, we need a solution that efficiently handles resources in space, frequency, and device dimensions. Semantic coordination could alternatively be used in such networks [135,164].

The nodes in the system can communicate and share knowledge in terms of the spectrum utilization in the network. In [164], the authors proposed to model the spectrum usage coordination as an interactive process between a number of distributed communicating agents, where agents share their specific information and knowledge. The information includes the current spectrum usage state, spatial coordinates of the device, available communication protocols, usage policy, spectrum sensing capabilities of the device, spectrum needs, etc. An approach for such coordination is presented in [164] and it is based on semantic technologies and communication between heterogeneous agents with potentially different capabilities and a minimal common compliance. The core knowledge is represented by ontologies whose representation and usage is specified in a standardized way. The approach is used as dynamic spectrum coordination algorithms used for coordination among different wireless technologies in 5G networking [135,164]. This semantic technologies based approach can be used for wide diapason of problems within 5G heterogeneous networks, such as network states predictions, network analysis, minimizing traffic load, content distribution coordination etc. This approach could be used in combination with caching techniques in order to improve content distribution in 5G, but further research should be done in this area.

**Mobile CDNs and the Cloud.** Mobile computing (MC) [13] has created enormous demand for online experience, that OSN-aware CDNs are required to satisfy. Almost-ubiquitous Wi-Fi coverage and rapid extension of mobile-broadband provide uninterrupted connectivity for mobile devices, whereas devices that hop seamlessly from WiFi to cellular networks, and technologies such as 5G, will be optimised for uses that put a premium on continuous connectivity regardless of the user location [5]. Mobile-specific optimizations for applications along with drastically simplified and more intuitive use of devices (e.g. with multi-touch interactions instead of physical keyboards) contribute to mobile applications becoming the premium mode of accessing the Internet, at least in the US [6].

Cellular networks have become the main way citizens connect to the Internet worldwide, specially in developing countries. Thanks to the development of mobile devices and their networking capacities, as well as the arrival of fast and reliable networks such as 5G, a high quality connectivity is ensured everywhere

and any time. The irruption of new paradigms, such as IoT, has increased the number of connected devices (sensors, actuators, etc.) which requires infrastructures that provide higher throughput networking, specially in use cases where high definition videos are involved and even new scenarios are yet to emerge.

Mobile Computing entails the processing and transmission of data over a medium, that does not constraint the human-medium interaction to a specific location or a fixed physical link. Figure 3 depicts a general overview of the MC paradigm in its current form. It is the present decade that signifies the proliferation of MC around the world, although handheld devices have been widely used for around two decades in the form of Personal Digital Assistants (PDAs) and early smartphones. Almost ubiquitous Wi-Fi coverage and rapid extension of mobile-broadband (around 78 active subscriptions per 100 inhabitants in Europe and America) provide undisrupted connectivity for mobile devices, whereas 97% of the world's population is reported to own a cellular subscription in 2015 [5]. Moreover, the MC paradigm is nowadays further combined with other predominant technology schemes leading to the paradigms of Mobile Cloud Computing [15], Mobile Edge Computing [8], Anticipatory Mobile Computing [138], etc.

Today's mobile devices include smartphones, wearables, carputers, tablet PCs, and e-readers. They are not considered as mere communication devices, as they are in their majority equipped with sensors that can monitor a user's location, activity and social context. Thus, they foster the collection of Big Data by allowing the recording and extension of the human senses [115].

Mobile social networking involves the interactions between users with similar interests or objectives through their mobile devices within virtual social networks [44]. Recommendation of interesting groups based on common geo-social patterns, display of geo-tagged multimedia content associated to nearby places, as well as automatic exchange of data among mobile devices by inferring trust from social relationships are among the possible mobile social applications benefiting from real-time location and place information.

1. *Industrial Applications*: Maintenance, service, optimization of distributed plant operations is achieved through several distributed control points, so that risk is reduced and the reliability of massive industrial systems is improved [139].
2. *Automotive Applications*: Automotive applications capture data from sensors embedded in the road that cooperate with car-based sensors. They aim at weather adaptive lighting in street lights, monitoring of parking spaces availability, promotion of hands-free driving, as well as accident avoidance through warning messages and diversions according to climate conditions and traffic congestion. Applications can promote massive vehicle data recording (stolen vehicle recovery, automatic crash notification, etc.) [7].
3. *Retail Applications*: Retail applications include, among many others, the monitoring of storage conditions along the supply chain, the automation of restocking process, as well as advising according to customer habits and preferences [139].



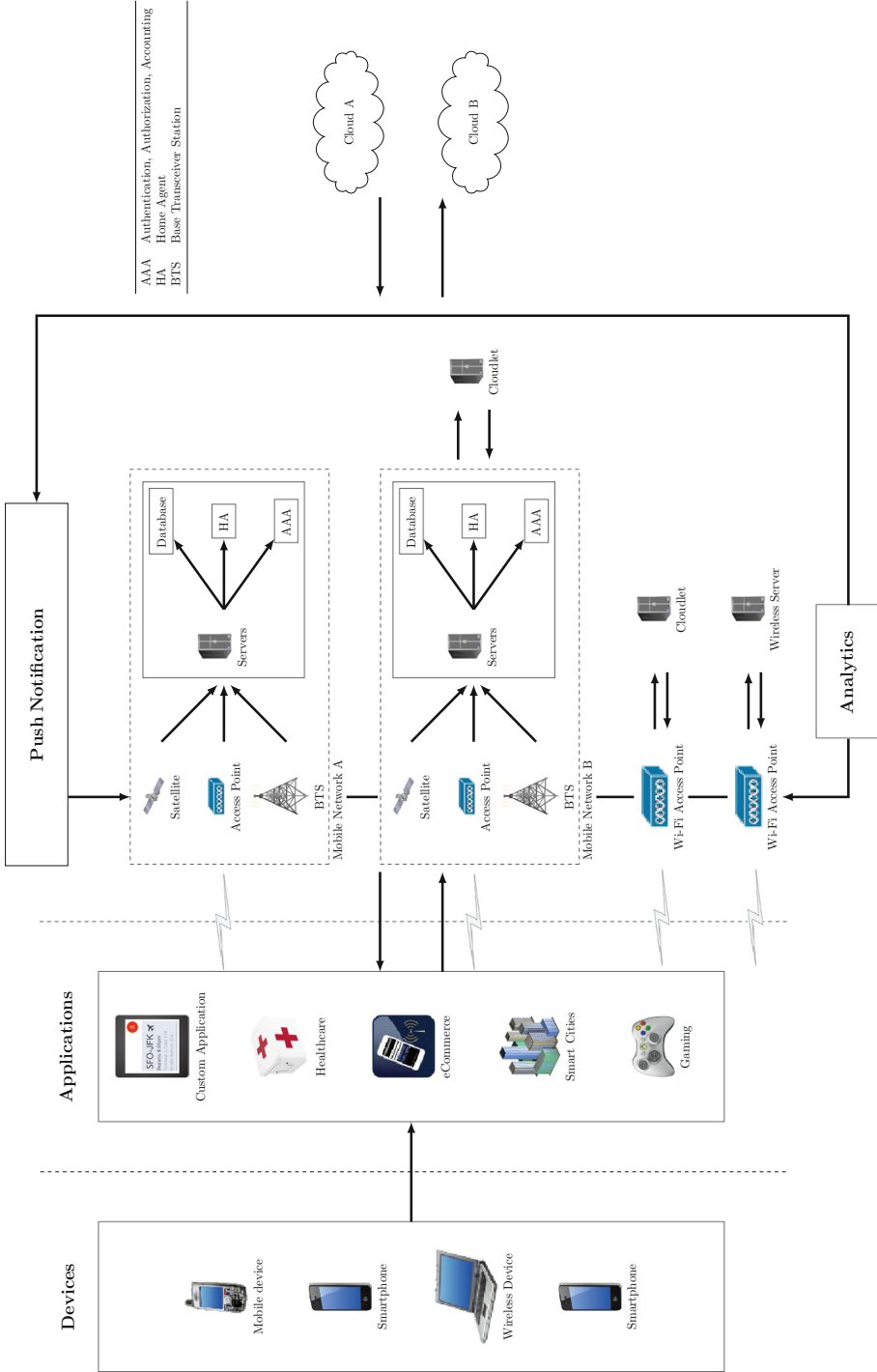


Fig. 3. Social data streams over mobile computing [108].

4. *Healthcare & Telemedicine Applications*: Physical condition monitoring for patients and the elderly, control of conditions inside freezers storing vaccines, medicines and organic elements, as well as more convenient access for people in remote locations with usage of telemedicine stations [98].
5. *Building Management Applications*: Video surveillance, monitoring of energy usage and building security, optimization of space in conference rooms and workdesks [7].
6. *Energy Applications*: Applications that utilize assets, optimize processes and reduce risks in the energy supply chain. Energy consumption monitoring and management [10,172], monitoring and optimization of performance in solar energy plants [167].
7. *Smart homes & Cities Applications*: Monitoring of vibrations and material conditions in buildings, bridges and historical monuments, urban noise monitoring, measuring of electromagnetic fields, monitoring of vehicles and pedestrian numbers to optimize driving and walking routes, waste management [81].
8. *Embedded Mobile Applications*: Applications for recommendation of interesting groups based on common geo-social patterns, infotainment, and automatic exchange of data among mobile devices by inferring trust from social relationships. Visual effects streaming workflow will give users on-demand, cloud-based access to visual effects tools, that can be accessed via web, given enough low-latency bandwidth to maintain a connection for streaming the User Interface from the cloud. Video Game streaming workflow will give players the option of streaming graphically-rich content that requires near-instant interaction between the game controller and the graphics on the TV screen [139].
9. *Technology Applications*: Hardware manufacture, among many others, is improved by applications measuring performance and predicting maintenance needs of the hardware production chain [139].

**Roadmap:** Our chapter is organized as follows: Sect. 2 discusses existent surveys concerning modelling, simulation and performance evaluation in the examined bibliographical field. The association of context-aware computing with social networks is given in Sect. 3. Infrastructure efficiency of deployed data centers for the distribution of social content is analyzed in Sect. 4 in terms of software solutions, as well as data center scheduling frameworks. Section 5 presets a categorization of most predominant models for the depiction of the information diffusion process in a social network. Section 6 discusses various architectures, systems and techniques for efficient content distribution based on social data streams, along with diverse studies that corroborate them as well as the way 5G network infrastructure affects the social data streams. Section 7 concludes and finally gives the outline of future research directions.

## 2 Related Work

In a manner that resembles the utilization of social data streams Anjum et al. [21] review the deployment of peer-assisted content delivery solutions. They present challenges caused due to heterogeneity in user access patterns and the variety of contextual information, such as interests and incentives of Internet Service Providers, End-Users and Content Providers. Furthermore, Perera et al. [139] survey context awareness from an IoT perspective. They indicate that the technology in the IoT is expected to enable expansion of conventional content delivery systems to a broader network of connected devices. They systematize the collection, modeling, reasoning, and distribution of context in relation to sensor data in a work that resembles the social data harvesting in terms of volume, variety and velocity. The survey also addresses a broad range of methods, models, systems, applications, and middleware solutions related to context awareness in the realm of IoT, that could be potentially applicable to social data streams, too.

In [111] Kilanioti et al. study various experiments on a modified content delivery simulation framework and compare miscellaneous policies for dynamic content delivery based on analysis of social data streams. The incorporation of an OSN-aware dynamic mechanism becomes indispensable for content delivery services, since (i) significantly large proportion of Internet traffic results from bandwidth-intensive multimedia content, that is produced via online media services and transmitted over OSNs, and (ii) multimedia content providers, such as YouTube, often rely on ubiquitous content distribution infrastructures. The policies presented take patterns of user activity over OSNs and exploit geo-social properties of users participating in extensive retransmissions of items over OSNs. The authors proceed to incorporate diverse caching schemes of the underlying infrastructure, miscellaneous policies for the handling of OSN data and various approaches that take into account the most efficient timing for content placement. The simulation framework introduced in [107] serves in this study as the basis of further parameterized content delivery experimentation that exploits information transmission over OSNs and decreases replication costs by selectively copying items to locations where items are bound to be consumed.

Downloads of large size multimedia contents are explored through several studies together with techniques that try to reduce doubled content transmissions using intelligent caching strategies in mobile networking [14,23,101]. The main idea is redistribution of mobile multimedia traffic in order to eliminate duplicated downloads of popular contents. Intelligent caching strategies would enable access to popular contents from caches of nearby nodes of a mobile network operator. Those strategies allow content providers to reduce access delays to the requested content. Many caching algorithms for content distribution already exist [23]. Efficient caching strategy could enhance the energy efficiency of 5G networks, thus the cooperative caching architecture is presented in [101]. This strategy addressed the increasing demand for mobile multimedia and data services in energy efficiency in emerging 5G systems using content caching and distribution.

We are not aware of surveys in the bibliography suggesting an holistic approach for the utilization of social data streams towards facilitation of content distribution decisions and social analysis tasks other than [108]. The diverse parameters we review in this work (modelling, simulation, performance evaluation) take into account low-level decisions and high-level considerations, including energy efficiency of employed data centers, in-memory keeping solutions and various network approaches for time-critical applications. We review combined aspects such as optimal route selection, data redundancy, data localization and data center optimizations.

### 3 Social Networks and Context-Aware Computing

A social network is a network of social bindings between people. Computer-Supported Cooperative Work (CSCW) has contributed much in offering advanced collaborative systems for leveraging human connections and improving human interactions in workspace environments, but these systems mostly focus on business-driven interactions where connections among people tend to be formal and structured [43]. Recently however, social and computing disciplines focused specifically on the design of social-networking services, i.e. applications that support human social interactions and can be more informal.

The advancement of wireless networks, as well as mobile, context-aware and ubiquitous computing, enabled the improvement of social-networking services by enabling social encounters between proximate users with common interests in an anywhere and anytime fashion, as in Ubiquitous Computing systems [43]. Thus, there has been a shift of the application focus from virtual to physical social spaces using ubiquitous technologies [43]. This shift introduces a great number of possibilities, however it also introduces a number of challenges that are related to ubiquitous computing. While social-network systems for ubiquitous computing environments are an emerging trend in social computing, due to the fact that ubiquitous-computing environments are more dynamic and heterogeneous than Internet based environments, appropriate solutions and design guidelines are required to facilitate their ubiquitous aspect.

Ubiquitous Computing, first introduced in the nineties, refers to the shifting of the computing paradigm from the desktop Personal Computer (PC) to a more distributed and embedded form of computing [170]. Together with Pervasive Computing (for many these terms are synonymous), Ubiquitous Computing introduced the concept of “anywhere, anytime computing”, allowing users to interact with computers embedded in every-day objects in an “anywhere and anytime” manner. Ubiquitous Computing specifies also that the interaction of users with such devices must be straightforward to the degree that the user would not be aware of such an interaction. Thus, in order for ubiquitous and pervasiveness to be achieved, computers must disappear from the front-end, be embedded to common objects that humans use daily and provide computational and informational services without expecting from users to explicitly and consciously interact with them.

Challenges in Ubiquitous Computing can be categorized to (Want and Pering [170]): (i) power management issues: refers to how mobile devices deal with processing power and storage space and the kind of wireless technology to use in every given situation, (ii) limitations in connecting devices: this issue has to do with how all these small devices will be connected and managed, (iii) user interface issues: since Ubiquitous Computing demands for many different small-scale devices of various types of interfaces and displays of various sizes, the challenge in user interfaces lies in developing user friendly and interactive interfaces to the level where users will be motivated in using them, (iv) issues related to Location Aware Computing. Henricksen et al. [88] add to the above list the challenge of managing heterogeneous devices of different hardware and software specifications, such as sensors and actuators, embedded devices in objects such as shoes, home and office appliances such as videos, mobile devices and traditional desktop computers, in order for these devices to interact seamlessly. Another challenge they mention has to do with maintaining network connections while devices move between networks of different nature and characteristics. In ubiquitous environments, people tend to use many devices simultaneously, therefore there is a need for these devices to communicate and exchange data. Another challenge Satyanarayanan [152] notes is tracking user intentions. This is important in Pervasive Computing in order for the system to understand what system actions could help the user and not hinder him/her.

An important challenge on context-awareness is to build context-aware systems that detect and manipulate the context in a human-like manner, i.e. making decisions proactively based on the context and provoke actions based on those decisions that assist the user through his/her task; the aforementioned should be done without any user participation or disturbance, except maybe in case of emergency. Another important issue is obtaining contextual information. Contextual information can be any information related to the user, the computing system, the environment of the user and any other relevant information regarding the interaction of the user and the system [63]. User's personal computing space can be used as the user's context (any information regarding the user taken from her personal profile, calendars, to-do lists etc.), various types of context can be sensed in real time like location, people and objects nearby, while contextual parameters could also include the current emotional and physiological state of the user. Contextual challenges also include the way context is represented (ontologies can be used or other context modeling techniques), the way this information is to be combined with the system information, as well as how frequently should context information be considered. Hinze and Buchanan [90] differentiate the static context from the fluent context. An example of static context is users profile information, while fluent context is dynamic, real-time context, e.g. time. The authors propose that a context model should be defined for each important entity, such as the user, the locations, etc. The authors mention as challenges the capturing of the context (whether it should be done automatically at particular times or manually by the user) and the process of storing the context (whether it should be stored on the client, on the server or both). On the

process of accessing contextual information, Hinze and Buchanan propose that context-awareness can help in reducing the amount of data to be accessed in real time, by pre-retrieving any relevant pre-known data, e.g. the static context [90], to increase efficiency.

User modelling is another challenge in developing ubiquitous systems. User modeling in ubiquitous environments is challenging: a user often changes roles depending on the context and the current environment he acts into; the big challenge is how to capture these changes and how to react on them [90].

Perhaps one of the most important contextual parameters is location, as it plays an important role in context-aware systems and ubiquitous systems. An issue with location as a contextual parameter is the type of location sensing technology to be used, while privacy is another issue. The issue with privacy is whether user privacy should be sacrificed for location awareness and to what extent. A third issue is the semantic (and contextual) representation of the location in order to utilize more contextual parameters than just the location itself. For example, by semantically representing locations, one can attach to them various information resources such as a webpage, a user profile, various objects with semantic representation etc. Schilit et al. [154], proposed the movement from the simplified concept of location to more contextually rich notions of place where people and activities should also be considered. Possible problems towards this concept include the difficulty in managing large scale positioning data, privacy concerns regarding location-awareness and the challenge of how to associate information objects, such as a web page, with a real-world location. Privacy issues regarding location-awareness are related to human psychology: users often consider privacy issues when their location is to be known by a system, but at the same time they provide private information such as credit card numbers and addresses to online systems without hesitation. This happens because in the first case they simply do not see the benefit of providing their location to be used by a simple application (e.g. finding friends in the proximity), while at the latter case they clearly see the benefit of buying goods online. The authors also argue that the centralized nature of the most location tracking applications (having a central server on which all user personal data are stored) discourages users from providing any personalized information, because centralized data can be accessed by anyone, not only illegally (e.g. hackers) but also the government, corporations with interest in user data (e.g. advertisers) etc. A solution can be the use of a decentralized schema where any personal data is stored and calculated on the client side, i.e. the user's device. An example of such a technology is the well known Global Positioning System (GPS): the client device uses satellite links to calculate locally the user's current position.

Context-Awareness and Adaptation related challenges and issues include

1. **Modelling the context:** which method is more appropriate to use
2. **Observing the context:** automatically or manually
3. **Context sensing:** how are contextual parameters retrieved (sensors, user profiles etc.). In retrieving context data from various sources (e.g. sensors), how are inconsistencies between these data resolved

4. **Accuracy of contextual information** should be well known during the design of ubiquitous systems
5. **Storing the context:** on server (privacy issues), on client or on both
6. **Accessing the context**
7. **Using the context**
8. **How are the user and the environment connected and interact**
9. **How will the application modify its behaviour (be adapted) based on the context**
10. **Systems should be more context-aware than just the location.** A place is more than a location (also a Location related challenge)
11. **Devices should not operate based only on their own context, but based on the context of the whole system**
12. **Contextual information should be used to reduce the amount of input that is needed from users** (also a Human-Computer Interaction related challenge)
13. **How to capture changes in the user's role** deals with capturing the current context (i.e. the environment and the various circumstances) and user modelling (what possible role could a person play according to context)
14. **Context should be processed and various components should adapt to it without interfering with user's task** – no user explicit interaction should be necessary
15. **Adaptation in ubiquitous environments:** may need to adopt various devices separately and at the same time, while the user maintains a consistent view for the system/application

Context-aware computing has evolved over time from desktop applications, web applications, mobile computing, pervasive/ubiquitous computing to IoT over the last decade [139]. Context-aware computing became more popular with the introduction of the term ‘ubiquitous computing’ by Mark Weiser, while the term ‘context-aware’ was first used by Schilit and Theimer [155] in 1994. Context-aware computing has proven to be successful in understanding sensor data. Advances in sensor technology led to more powerful, cheaper and smaller sensors. The number of employed sensors is expected to grow over the next decade [161], generating ultimately big data [139, 140].

In settings where social communities become mobile, i.e. users not only interact, meet and communicate via social networks, but are mobile as well (move into the environment, interact with others, etc.), the concept of group awareness is met [53, 134, 176] where context related to the group is exploited to enable ubiquitous applications and services to function and serve people's concerns and needs in a pervasive manner. There is a need, thus, for formulating dynamic communities aiming to facilitate people in performing common tasks. It is often the case that such dynamic communities are resolved after the current goals have been achieved [134]. It is evident, thus, that the context within which such dynamic communities are created, act, achieve goals and are then resolved is important, and that, through this context, we can understand the groups' interests and, thus, personalize the applications and services offered [134].

A bibliography study [134] on mobile social network applications and platforms states that the context features that these applications and platforms use can be summarized as follows: Location, Interest, Time, Personal, Activity and Social Interaction. Here, context is “any information that can be used to characterize the situation of an entity” [62] and social context is “the information relevant to the characterization of a situation that influences the interactions of one user with one or more other users” [169]. Moreover, in [134] a context-aware Mobile Social Network model is proposed aiming to facilitate the creation of dynamic social networks based on a combination of multiple contexts, including location, users’ profile, domain specific data and OSN data, along with services for fostering the interaction among users.

## 4 Infrastructure Efficiency

### 4.1 Software Solutions for Infrastructure Efficiency

Regarding infrastructure efficiency, various models have been proposed for the optimization of such infrastructures that support social networks data centers. These approaches have also been proposed by industry partners addressing various challenges [42, 129, 163]. Among these challenges, the following have been identified: (a) data volume increase, (b) confinement, (c) interactions made in real-time or near real time, (d) federation of data center infrastructures and (e) cdn-load-balancing between data centers, but usually not focused on cost effectiveness. One of the main data center costs is energy consumption of both the IT equipment as well as the supporting Mechanical and Electrical (M&E) infrastructure. A widely used indicator that measures the energy effectiveness of the M&E infrastructure overhead is Power Usage Effectiveness (PUE) [25], which is calculated as the *Total facility energy/Total IT energy* and has a theoretical minimum of 1.

Figure 2 shows the categories on which the research community and other stakeholders have developed solutions for the improvement of costs and efficiency:

- **Workload Consolidation and Shut-down Techniques,**
- **VM Consolidation and Migration,**
- **Data Replication and Placement, and**
- **Energy-aware Scheduling Algorithms.**

The main objective based on these solutions is to reduce the idleness of computing and storage nodes (throttle resources), while switching off unused machines without jeopardizing Service Level Agreements. Some representative examples from each category are shown in Table 1.

Regarding **Workload Consolidation and Shut-down Techniques**, heuristics for energy-aware consolidation of jobs and maximization of resource utilization are presented in [117]. These approaches estimate resource consumption in terms of CPU utilized by tasks and encourage resources to execute multiple tasks in parallel. The proposal from [102] is an algorithm that search for a



minimum multi-objective function, taking into account energy-consumption and running time by combining resource allocation and heuristic rules and simulating Directed Acyclic Graph (DAG) based workloads.

One of the most popular approaches to increase efficiency is to switch off idle servers [75, 78], which is usually tested in dedicated simulation tools [74]. Several models, including games theory models [72, 73], are used to balance opposite requirements in data centers, such as performance and energy consumption. Even models utilized for economic environments, such as Data Envelopment Analysis, are employed to analyze the efficiency in various realistic large-scale data centers and propose corrections to improve data-center efficiency [76].

In addition, techniques for energy conservation like **Virtual Machine (VM) Migration and Consolidation** are widely studied and already employed. In [35], a resource manager solution focused on virtualized data centers, which enables lower energy consumption by applying VM migrations and allocations based on current CPU usage, is proposed. An extension of VM migration is presented in [34] where Service Level Agreement (SLA) restrictions are considered. Allocation and migration of VMs is also the target in [159] where a Bayesian Belief network algorithm is presented. Moreover, a day/night pattern is taken into account for an energy manager in [148], based on the aggregation of traffic during low usage periods and shutting down idle machines.

Solutions for improving energy proportionality through **Data Replication and Placement**, are also available. A power-proportional distributed file system approach that tries to store data on non-overlapping subsets of machines is presented in [18]. Such subsets of machines contain only one copy of each file and administrators can decide how many subsets will be turned on to serve incoming requests. On the other hand, a division of the cluster in non-overlapped zones is proposed in [162], enabling operators to shut down zones. In a similar way, in [105] the authors present a variation of the Hadoop File System (HDFS) that divides the cluster in *Hot Zones* that store recent data and *Cold Zones* where low [e.g. spatial or temporal] popularity files are stored. Then a power off policy is applied to the *Cold Zones*. A non-uniform replica placement on data popularity is also presented in [125].

**Energy-Aware Scheduling Algorithms** is the last family of solutions. A green scheduling algorithm based on neural networks is proposed by [68], focusing on the prediction of workload demand with the purpose of applying power-off policies to idle servers. Experiments presented simulate a medium sized data center that runs homogeneous workload that is intended to respond to end-user requests. Energy-aware scheduling policies combined with Dynamic Voltage and Frequency Scaling (DVFS) is presented in [100]. In [77], a multi-objective scheduling algorithm is proposed, based on genetic algorithms, which takes into account energy efficiency, performance and security constraints.

We have classified the related work under consideration in terms of their final objective, including: (a) modelling, (b) simulation, (c) performance. Such classification is shown in Table 2.

**Table 1.** Related work summary

|              |   |                     |
|--------------|---|---------------------|
| Ref<br>[68]  | Title: Performance evaluation of a green scheduling algorithm for energy savings in cloud computing                                   | Savings<br>~ 45%    |
|              | Category: Workload consolidation and power off policies (power off policy based on a neural network predictor)                        |                     |
|              | Evaluation: [8-512] nodes cluster simulation  |                     |
|              | Workload: End user homogeneous requests that follow a day/night pattern   |                     |
| Ref<br>[117] | Title: Energy efficient utilization of resources in cloud computing systems   | Savings<br>[5-30]%  |
|              | Category: Workload consolidation and power off policies (energy-aware task consolidation heuristic based on different cost functions) |                     |
|              | Evaluation: Simulation of a not stated size cluster   |                     |
|              | Workload: Synthetic workload in terms of number of tasks, inter arrival time and resource usage                                       |                     |
| Ref<br>[148] | Title: Saving energy in data center infrastructures   | Savings<br>[20-70]% |
|              | Category: Workload consolidation and shut-down techniques (safety margin power-off policy)  |                     |
|              | Evaluation: 100 and 5000 nodes cluster simulation   |                     |
|              | Workload: Synthetic workload that follows a day/night pattern   |                     |
| Ref<br>[35]  | Title: Energy efficient resource management in virtualized cloud data centers   | Savings<br>~ 80%    |
|              | Category: VM consolidation and migration  |                     |
|              | Evaluation: 100 nodes cluster simulation using CloudSim   |                     |
|              | Workload: Synthetic workload that simulates services that fulfill the capacity of the cluster   |                     |
| Ref<br>[102] | Title: dynamic energy-aware scheduling for parallel task-based application in cloud computing   | Savings<br>[20-30]% |
|              | Category: Energy-aware scheduling algorithms (polynomial-time and multi-objective scheduling algorithm for DAG jobs)                  |                     |
|              | Evaluation: Experimentation on a 64 nodes cluster   |                     |
|              | Workload: Synthetic directed acyclic graph-based workload   |                     |

## 4.2 Data Center Scheduling Frameworks

Resource managers have direct impact on the efficiency of the infrastructure since they are responsible for the application of energy-aware scheduling models. The responsibility for actually deciding resource negotiation and tasks deployment have range from traditional approaches to fully managed solutions such as data centers which are used by many entities with multiple users and various kind of applications and requirements [30, 38].

**Table 2.** Classification of approaches according to their objective. The green color represents that the work focuses strongly on that objective, and the red color represents opposite.

| Reference | Modelling | Simulation | Performance |
|-----------|-----------|------------|-------------|
| [101]     | ●         | ●          | ●           |
| [103]     | ●         | ●          | ●           |
| [118]     | ●         | ●          | ●           |
| [34]      | ●         | ●          | ●           |
| [36]      | ●         | ●          | ●           |
| [160]     | ●         | ●          | ●           |
| [149]     | ●         | ●          | ●           |
| [69]      | ●         | ●          | ●           |
| [18]      | ●         | ●          | ●           |
| [106]     | ●         | ●          | ●           |
| [126]     | ●         | ●          | ●           |
| [163]     | ●         | ●          | ●           |

The constraints imposed by diverse applications in terms of size, inter-arrival and duration, may lead to various resource efficiency, latency rates and security levels.

Current trends aim to utilize the same hardware resources to deploy various kind of applications and frameworks with diverse requirements, which increases the complexity since diverse data are to be processed.

We present the main categories of the scheduling resource-managing models following several approaches and we show their limitations summarized in Table 3.

Monolithic models, where a centralized manager is responsible for all scheduling and resource managing decisions came first. Such models [96] are a good choice when the workload is composed of a relative low number of Batch jobs, due to these schedulers being omniscient [55], since such kind of workload does not usually have strict latency requirements [60]. Monolithic resource managers perform near-optimal scheduling operations [58, 85, 175] as they are able to completely examine the data center. This detailed inspection allows the determination in terms of performance implications and impact on shared resources. [83, 128, 133, 158, 173]. Due to this detailed cluster inspection, monolithic centralized schedulers usually utilize resources at a higher level than other approaches [168]. Monolithic centralized schedulers also achieve high-quality decisions which result in shorter makespans, near-optimal load balancing and predictable performance and availability [59, 156, 177].

**Table 3.** Cluster scheduling approaches.

| Frameworks                   | Strategy                    | Optimal environment   | Near-optimal environments  |
|------------------------------|-----------------------------|---|--|
| Paragon [58]                 | Centralized                 | Low number of long-running and non-latency sensitive jobs   | Mid and high number of jobs  |
| Quasar [59]<br>Borg [168]    | Monolithic                  |   | Mixed workloads  |
| YARN [166]                   | Centralized<br>Two-level    | Mid number of diverse long-running, non latency sensitive jobs  | Latency-sensitive workloads  |
| Mesos [89]                   |                             |   |  |
| Omega [156]                  | Centralized<br>Shared-state | Mid number of heterogeneous workloads   | High number of short and latency-sensitive jobs                        |
| Canary [142]                 | Distributed                 | High number of short, non resource-demanding, latency-sensitive jobs  | Long, resource-demanding jobs  |
| Tarcil [60]<br>Sparrow [137] |                             |   | Mixed workloads  |
| Mercury [103]                | Hybrid                      | Mixed workloads composed of 90% of short, latency-sensitive jobs and 10% of long-running, resource-demanding jobs | Homogeneous workloads<br>Other workloads patterns<br>Evolving patterns |
| Hawk [57]<br>Eagle [56]      |                             |   |  |

With the arrival of new computation paradigms such as microservices, current trends tend to divide jobs into smaller parts which usually are more latency-sensitive. This new scenario with huge amounts of small jobs overcome the capacity of Monolithic models. Two new centralized resource managing models were proposed to overcome this limitation by dividing the responsibility of resource managing and scheduling:

- Two-level resource managers, such as Mesos [89], and YARN [166] employ a central resource manager which coordinates a set of independent schedulers. The parallel schedulers pessimistically block the data center in order to make a scheduling decision. Such manager offers resource schedulers, and as a response, the set of schedulers perform scheduling decisions for deciding which machines will execute a particular task. The down side of this model is that, opposed to Monolithic models, the schedulers are not omniscient. In this model, data-center state and tasks requirements are not always available to make optimal scheduling decisions.
- Shared-state resource managers, such as Omega [156], employ a centralized manager which orchestrates a set of parallel scheduling agents. In contrast to Two-level resource managers, each scheduling agent makes scheduling decisions based on a partially out-of-date copy of the whole data-center state.

Instead of blocking the data center to apply their decisions, they follow a transactional approach. If a transaction ends up in a conflict, the data-center state is requested and the scheduling restarts.

However, all the aforementioned proposals suffer from a performance bottlenecks when huge latency-sensitive workloads composed of millions of tasks are under consideration, as they employ centralized coordinators for resource managing or even for resource managing as well as scheduling [137].

Distributed schedulers such as Sparrow and Canary [67, 137, 142, 143] are built to work optimally when the aforementioned scenarios are considered. Distributed models employ faster and simpler algorithms in order to analyze smaller areas of the data center, which leads to sub-optimal decisions with higher throughput and lower latency rates.

The frameworks and applications served by data centers are constantly evolving. Current trends show that in most cases heterogeneous workloads are being deployed in large-scale realistic clusters [50, 145]. Such workloads are composed by two main kinds of jobs: (a) Jobs such as web servers and data-center frameworks, which represent 10% of jobs. These jobs consume, however, more than 80% of computing resources because they run for long periods; and (b) Jobs such as MapReduce tasks, which represent 90% of jobs. These jobs run for shorter periods and consume less than 20% of computing resources [19, 144, 147, 175]. In such environment, sub-optimal scheduling operations may severely impact on the aforementioned large jobs. Hence, distributed models may achieve worse results in terms of performance compared to those achieved by centralized models.

Finally, hybrid models, such as Hawk and Mercury [56, 57, 103] were developed to work well under the aforementioned scenario. These models employ centralized and distributed approaches in order to overcome the limitations discussed. Hybrid models use a centralized scheduler for long-running jobs to provide high-quality scheduling and, on the other hand, they employ a distributed approach for those short jobs which need quick scheduling to achieve latency goals.

Beyond workload consolidating and resource throttling, other research has explored the impact IT hardware refresh and optimization could have on data centre energy consumption [31]. Additionally, it was shown that addressing energy efficiency at the design stage of software systems presents a significant opportunity to reduce infrastructure energy consumption [32].

## 5 Content Diffusion Models for Social Data Streams

This section outlines the most predominant models for the depiction of the content/information diffusion process in a social network described in [108]. Most of the existent algorithmic solutions for content distribution are built on them, thus the assumption that content circulation over social data streams is depicted by one of them is of crucial importance for the suggested solutions. The main algorithmic problems studied in the bibliography are related with the discovery of nodes that are most prone to diffuse content to the greatest extent, and the

categorization of nodes according to their influence degree. The categorization of the models is depicted in Fig. 4. The models presented are the most recent in the bibliography and there are no prior recent models to the best of our knowledge. The first-level discrimination of models is based on whether they take the structure of the network into consideration (*network-aware*) or not (*holistic*). In other words the discrimination criterion is if they incorporate knowledge about underlying associations of the nodes (edges) or, to the contrary, follow an aggregate-level approach.

*Information Diffusion Models Classification.* Li et al. in [120] classify information diffusion issues as a “3W issue”, that is with regard to “What”, “Why” and “Where”. They consider “What” to refer to the question “what latent information is there to be found in social networks?” and they provide as an example the findings, such as the way that an individual’s shopping habits relate to his/her profession, that can be included in a large volume of consumer data. Furthermore, their proposal considers “Why”, to refer to the question “why has the information propagated in this way?” This question refers to the factors that have affected the diffusion result, for example the factors that have produced a particular social cascade. Finally in their view “Where”, refers to the question “where will the information be diffused to in the future?”. This question refers to the future diffusion path that will be followed. For example if two influential users receive the same information from a common contact in a social network, but have a different perspective on the information, then it is important to estimate how they will respond and whether they will propagate the information through the network.

Based on the “3W issue” they classify information diffusion models as predictive and explanatory. Explanatory models aim to discover answers to important questions concerning the information diffusion process, such the determination of the main factors that affect information diffusion and the most influential nodes in the network. Predictive models, on the other hand, are used to predict the future information diffusion process in social networks based on certain factors, for example the quality of information diffused.

In a similar manner Luu et al. in [126] classify information diffusion models in non-network and network diffusion models. The former refers to user communities without any knowledge about the user relationship network and the latter is more applicable to the social networks where user relationships networks are given (e.g. Facebook, blog networks). For each model category Luu et al. describe representative models, notable model extensions, as well as model applications. The surveyed applications include Influence Maximization and Contamination Minimization. Model extensions are asynchronous models that incorporate time delay factors into the basic models.

An important class of information diffusion models are inspired from natural and biological systems. A typical example is provided by Dewi and Kim [61], who propose a bio-inspired model for information diffusion in complex networks using ant colony optimization. The model introduces selfishness in forwarder nodes and unacquainted nodes and employs ant colony optimization to find shortest path

and manage the selfish nodes and disjointed nodes. The authors provide simulation results in two types of networks: lattice networks and scale free networks, and the results show that the ant-colony model has higher performance and higher reachability than a selected baseline epidemic model.

*Influence Maximization.* Influence maximization is an important issue in social network analysis domain which concerns finding the most influential nodes in a social network. Determining the influential nodes is made with respect to information diffusion models and is based on the observation that most of the existing models only contain trust relationships. In this respect Hosseini-Pozveh et al. in [93] classify influence maximization models in two classes: cascade-based and threshold-based. They evaluate all models in comparison with selected benchmark models through two real data sets, the Epinions and Bitcoin OTC. Based on the evaluation results a main conclusion is drawn: when a distrusted user performs an action or adopts an opinion, the target users may tend not to do it.

The efficiency of influence maximization algorithms is subject to active research since the problem is known to be NP-hard. In this respect Kempe et al. proposed a greedy algorithm (referred to as SimpleGreedy) that guarantees 63% influence spread of its optimal solution [106]. Along this line Ko et al. in [113] propose an improved algorithm, termed Hybrid-IM, which by combines PBIM (Path Based Influence Maximization) and CB-IM (Community Based Influence Maximization). Ko et al. further provide evaluation results from extensive experiments with four real-world datasets. They show that Hybrid-IM achieves great improvement (up to 43 times) in performance over state-of-the-art methods and finds the seed set that provides the influence spread very close to that of the state-of-the-art methods.

## 5.1 Holistic View Models

Rogers' theory [150] is quantified by the Bass model [33]. The Bass model is based on the notion that "the probability of adopting by those who have not yet adopted is a linear function of those who had previously adopted" (F.Bass). It predicts the number of adopters  $n(t) \in N$  of an innovation at time  $t$  (in the information diffusion scenario the number of retransmitters of an information piece):

$$n(t) = pM + (q - p)N(t) - q/M ( N(t) )^2 \quad (1)$$

where  $N(t)$  is the cumulative number of adopters by time  $t$ ,  $M$  is the potential market (the ultimate number of adopters),  $p \in [0, 1]$  is the coefficient of innovation (the external influences, expressing the individuals influenced by the mass media), and  $q$  is the coefficient of imitation (internal influence, expressing the individuals influenced by the early adopters). This approach, however, largely ignores the underlying network structure.

Models under the same concept of holistic view of the social behaviour make use of differential equations, and include, among others, the “multi-step flow model” by Katz and Lazarsfeld [104], the Daley-Kendall rumours model [54], and also, more recent ones, such as, the Van den Bulte and Joshi model of influentials and imitators [47].

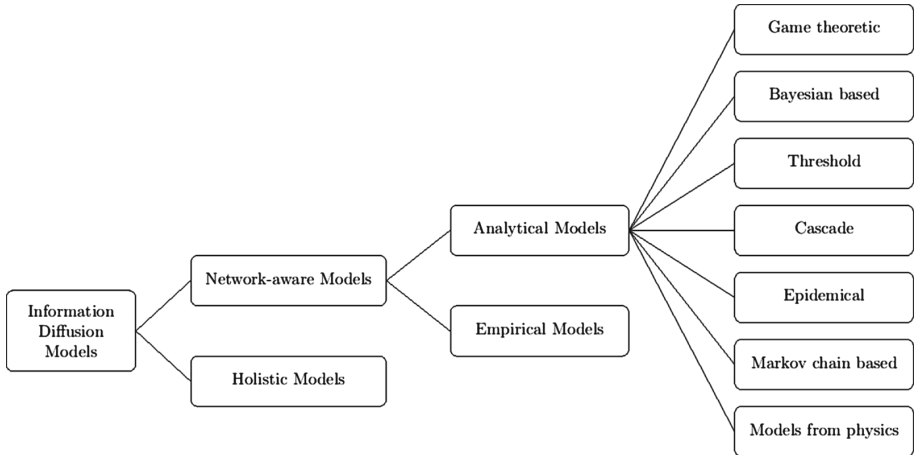


Fig. 4. Content/information diffusion models

## 5.2 Network-Aware Models

These include completely novel models, but also variations of the aforementioned (holistic) models, such as the Nekovee variation [136] of the Daley-Kendall model, and are separated in following categories, based on whether they are mathematically formulated (*Analytical models*) and then applied or are the outcome of empirical methods, such as regression, regression trees etc. (*Empirical models*).

**Analytical Models.** The first mathematical models based on nodes’ thresholds for the depiction of information diffusion were developed by Schelling [157] and Granovetter [86]. A categorization of the most predominant models is presented.

*Game-Theoretic Models.* In [112], Kleinberg proposes a simple networked coordination games model. The author assumes that there are two behaviours a node  $v \in V$  in the graph  $G = (V, E)$  can follow, A and B. The model is based on the notion that for each individual the benefits of adopting a new behaviour increase as more of its neighbours adopt the new behaviour. At discrete time steps each node updates its choice of A or B according to the behaviour of its neighbours. The objective of the nodes is to switch each time to the behaviour that reaps the maximum benefit for them. For the nodes  $v$  and  $w$  there is a motivation for



behaviour matching, expressed in the following way, where parameter  $q$  is a real number  $0 < q < 1$ :

- if  $v$  and  $w$  both choose behaviour A, they both receive a  $q$  payoff
- if  $v$  and  $w$  both choose behaviour B, they both receive a  $1 - q$  payoff
- if  $v$  and  $w$  choose different behaviours, they both receive a 0 payoff

$v$ 's payoff for choosing A is  $qd_v^A$  and for choosing B is  $(1 - q)d_v^B$ . The overall payoff for  $v$  playing the game with its neighbours in  $G$  is the sum of the individual (pairwise) payoffs;  $q$  is actually the threshold expressing the fraction of adopting neighbours, since it easily results that  $v$  should adopt behaviour B if  $d_v^B > qd_v$ , and A if  $d_v^B < qd_v$ , where  $d_v$  is the degree of the node,  $d_v^A$  the number of its neighbours with behaviour A and  $d_v^B$  the number of its neighbours with behaviour B.

Initially there is a set  $S$  of nodes adopting behaviour B and  $h_q(S)$  is the set of nodes adopting B after one round of updating with threshold  $q$ .  $h_q^k(S)$  is the set of nodes adopting B after  $k$  successive rounds. A set  $S$  is *contagious* (with respect to  $h_q$ ) if “a new behaviour originating at  $S$  eventually spreads to the full set of nodes” and the contagion threshold of a social network  $G$  is “the maximum  $q$  for which there exists a finite contagious set”.

The technical issue of *progressive* or *non-progressive processes* (monotonous or non-monotonous as referred to later on in the present study) refers to the fact that when a node  $v$  following till then the behaviour A updates to behaviour B in time step  $t$ , it will be following B in all subsequent time steps. Although, intuitively, we would expect progressive processes to give finite contagious sets more easily (because of lack of early adopters setbacks that would hinder the cascade), Kleinberg points out that both the progressive and non-progressive models have the same contagion thresholds [131], which in both cases is at most  $1/2$  (“a behaviour can't spread very far if it requires a strict majority of your friends to adopt it”) [131].

More game-theoretic models can be found in the work of Arthur [22], who proposes a simple cascade model of sequential decisions with positive externalities, manifested by a term that adds to the payoff of a decision. Namely in the scenario of two competing products, the latter become more valuable as they are used by more users (for a social media site or a smartphone, for example, it will acquire better third-party applications and support as its users grow). Also game-theoretic models are introduced by Banerjee [29] and Bikhchandani et al. [39], that are based on influence not due to positive externalities, but because of information conveyed from earlier decisions. The proposed game-theoretic models, however, have the drawback of not taking heterogeneity into consideration, in the notion that all nodes have the same threshold, and all their neighbours contribute the same in making a node change its behaviour.

*Bayes-Based Models.* Combining nodes' private information and their observations of earlier adoptions, in [69], Kleinberg and Easley present a Bayes based model to formulate information cascades, answering questions such as “What is

the probability this is the best restaurant given the reviews I have read and the crowds I see there?”.

$$Pr[A|B] = \frac{Pr[A] Pr[B|A]}{Pr[B]} \quad (2)$$

Three factors are taken into consideration:

- The states of the world;
- Payoffs; and
- Signals.

The first factor expresses whether an option is good or bad (if a new restaurant is a good or a bad choice). Supposing that the two options of the world are K (the option is a good idea) and B (the option is a bad idea), the world is placed in K with probability  $p$  and in B with probability  $1 - p$  ( $Pr[K] = p$ ,  $Pr[B] = 1 - Pr[K] = 1 - p$ ). Payoffs for a node  $v$  are defined as follows:

- If  $v$  rejects the option, the payoff is 0.
- If  $v$  adopts a good idea, it receives a positive  $v_g > 0$  payoff.
- If  $v$  adopts a bad idea, it receives a negative  $v_b < 0$  payoff.
- If  $v$  adopts without any prior knowledge, the payoff is 0.

The signals refer to private information each individual gets about the benefit or not of a decision: a high signal ( $H$ ) suggests that adoption is a good idea, whereas a low signal ( $L$ ) suggests that it is a bad idea. If accepting is indeed a good idea, then  $Pr[H|K] = q > \frac{1}{2}$  and  $Pr[H|B] = 1 - q < \frac{1}{2}$ . In the restaurant example the private information could be a review that an individual reads about the first restaurant, with a high signal corresponding to a review comparing it favorably to restaurant B. If choosing the first restaurant is indeed good, there should be a higher number of such reviews, so  $Pr[H|K] = q > \frac{1}{2}$ . Kleinberg and Easley [69] consider how individual decisions are made using (Eq. 2) when they get a sequence of independently generated signals consisting of a number of high signals and a number of low signals, thus making interesting observations about situations where individuals can observe others’ earlier decision, but do not have access to their knowledge.

The basic propagation models on which most generalizations for information diffusion are based are the Linear Threshold Model (LTM) [86, 157, 171] and the Independent Cascade Model (ICM) [82] with many proposed extensions (LTM: [106, 171], ICM: [82, 87, 106]) and also a proposed unification [106].

*Linear Threshold Model.* Based on the assumption that some node can be either active (adopts a new idea/transmits a piece of information) or inactive and taking into account the monotonicity assumption, namely that nodes can turn from inactive to active with the pass of time but not the opposite, we can say that the LTM is based on the following notion: Each node  $v$  has a predefined activation threshold  $\theta_v \in [0, 1]$ , which expresses how difficult it is for the node to be influenced when its neighbors are active (“the weighted fraction of the

neighbors of node that must become active in order for node to become active”), and is influenced by each one of its neighbors  $w$  according to a weight  $b_{vw}$ , so that  $\sum_{w \in \Gamma(v)} b_{vw} \leq 1$ . The thresholds can be produced randomly with a uniform distribution, but some approaches investigate a uniform threshold for all the nodes of the network, e.g. [37]. The process takes place in discrete steps and the nodes that satisfy the constraint  $\sum_{w \in \Gamma(v)} b_{vw} > \theta_v$  are gradually added as active to the initial set of nodes. It’s worth mentioning that LTM can result as a modification of the networked coordinations game referred in the previous paragraph with the differentiation of payoffs for different pairs of nodes.

LTM expresses the idea that the influence of the neighbours of a node is additive, but when the rule of influence can not be expressed by a simple weighed sum, for example a node becomes active when one of its acquaintances and two of its co-workers do so, the arbitrary function  $g_v$  substitutes the weighed sum. In the *General Threshold Model* for time steps  $t = 1, 2, 3, \dots$  a node  $v$  becomes active if the set of active neighbours at  $t$  satisfy  $g_v(X) > \theta_v$ .

*Independent Cascade Model.* Under the ICM model [82], there is also a set of initially active nodes, the process takes place in discrete steps, but when node  $v$  becomes active, it has only one chance of activating each of its inactive neighbors  $w$  until the end of the process with a probability  $p_{vw}$  independent of the activations history and with an arbitrary order.

Exact evaluation of activation probabilities is exponential to the number of edges of the graph. Improving the performance of the works in [86] and [160], there are works studying the calculation of these probabilities such as [84] (based on a General Threshold Model with the assumption that each parent’s influence is fixed), or [64] (based on the ICM). In the latter, sampling from the twitter dataset is conducted in an efficient Markov-Chain Monte Carlo fashion using the Metropolis-Hastings algorithm [51] and the problem is tackled with two differentiations, one of which considering the past paths of data known (retweets for the twitter dataset) and one considering only the past path endpoints known (hashtags and urls) and joint probabilities are taken into consideration, reflecting also model uncertainty.

*Epidemical Models.* In the case of epidemical models a single activated (“infected”) node causes the change of state of a neighbour susceptible node, whereas in the afore-mentioned threshold and game-theoretic models a node has to interact with multiple neighbour nodes to evolve (complex contagion).

Epidemical models were introduced on the assumption that information would propagate like diseases. They constitute another category with an almost straightforward pairing with the ICM. The ICM captures the notion of contagion more directly, and also allows us to incorporate the idea that a node’s receptiveness to influence does not depend on the past history of interactions with its neighbors.

Epidemical models variations include the simple branching processes model, where a node infects a number of nodes and the contagion proceeds in subsequent waves with a probability  $\pi$ . This model is characterized by the basic reproductive

number of the disease  $R_0 = k\pi$ , where  $k$  is the number of new people somebody meets, which expresses the anticipated number of new cases of the disease that a single node will cause.

Extensions of the epidemical models are the SIR, SIS, and SIRS models: S stands for susceptible nodes, nodes that have not been infected yet and have no immunity to the contagion. I stands for infected nodes, nodes contagious to their susceptible neighbours, and R stands for recovered nodes, with the recovery considered as permanent in SIR and temporary in the case of SIRS [114]. The sequence of the letters in the acronyms of the models explains the flow of the epidemic. In SIR model nodes pass from the state of being susceptible to the state of being infected and then recover. In SIS model nodes are immediately susceptible once they have recovered (like in the case of common cold, recovery does not imply immunity that lasts for long). In the SIRS model recovered nodes free of infection may rejoin the susceptible nodes.

*Markov Chain Models.* Markov chains [65] are used to describe transitions from one state of a system to another in a finite set of possible states. Their memoryless nature (*Markov property*) has to do with the fact that the next state each time is independent of the preceding states. More formally: With a set of states  $\{\xi_1, \xi_2, \dots, \xi_r\}$  the process moves successively from one state to another in so-called *steps*, and specifically from state  $\xi_i$  to state  $\xi_j$  with a probability  $p_{ij}$  (*transition probability*) independent of the previous states of the chains, or remains in the same state with a probability  $p_{ii}$ . A particular state is picked from  $\Xi$  as the initial state. Markov chains are usually depicted with a directed graph, where the edges' labels denote the transition probabilities.

Markov models are widely used for analysing the web navigation of users. PageRank [45] is based on a Markov model and is used for ranking of information in the World Wide Web. By assigning weights that denote the relative importance of an hyperlinked document in a set of documents, the likelihood that a person will reach a specific page through random clicks is, essentially, represented.

In [160], Song et al. use a Continuous-Time Markov Chain Model (CTMC), namely a Markov model that describes the transition among states after some time of stay in a particular state. This time is exponentially distributed and does not affect the transition probability to the next state. The information diffusion model is introduced on a network  $G(V, w, \tau)$ .  $G$  contains a set  $V$  of  $n$  nodes and  $E$  edges between nodes representing the information diffusion paths.  $w$  denotes the set of the edges' weights ("amount of information to flow from one node to another") and  $\tau$  the set of the time delay on the information diffusion paths. Thus, the representation of the graph matches the CTMC in the notion that each node represents a state, each weight a transition probability and the delay is represented as the time-to-stay in each state.

*Voter Model.* The basic voter model introduced by Clifford and Sudbury [52] and Holley and Liggett [91], is defined in an undirected network and allows the spread of two opinions. In discrete time steps, a node adopts the opinion of a

randomly chosen neighbour. For a node  $v \in V$  in graph  $G = (V, E)$ ,  $\Gamma(v)$  is the set of neighbors of  $v$  in  $G$  and initially the nodes are arbitrarily endowed with a 0/1 state. At time step  $t$  each node adopts the opinion of one uniformly picked neighbour. With an initial assignment  $f_0 : V \rightarrow \{0, 1\}$  inductively we define

$$f_{t+1}(v) = \begin{cases} 1, & \text{with probability } a \\ 0, & \text{with probability } b \end{cases} \quad (3)$$

where  $a = \frac{|\{u \in \Gamma(v) : f_t(u) = 1\}|}{|\Gamma(v)|}$  and  $b = \frac{|\{u \in \Gamma(v) : f_t(u) = 0\}|}{|\Gamma(v)|}$ .

Even-Dar and Shapira [70] argue that it is one of the most natural probabilistic models to capture the information diffusion in a social network. It is suitable for depicting the spread of a technological product, as it is proved that under this model consensus is reached with probability 1. Even-Dar and Shapira refer to the (almost) consensus of products such as Google as a search engine, YouTube as a video-sharing website etc.

*Models from Physics.* Models from physics include the Ising model [97] serving for the description of magnetic systems, and bootstrap percolation [16] serving for the description of magnetic systems, neuronal activity, glassy dynamics, etc.

The Ising model [97] was first proposed in statistical physics and encompasses the notion of a ground state (in physics the state with the minimum energy), and that of the “self-optimizing” nature of the network.

Similarly to the basic voter model, there can be two competing “opinions”, in favour of or against a subject, let’s say depicted by a “+1” and a “-1”, which in physics express the correspondence of an atom forming a network to a spin variable (can be considered as the basic unit of magnetization) state  $\sigma_i = \pm 1$ . The total energy of the system under this model (Hamiltonian) is defined as:

$$H = H(\sigma) = - \sum_{\langle i, j \rangle} E \sigma_i \sigma_j - \sum_i J \sigma_i \quad (4)$$

for each configuration  $\sigma = (\sigma_1, \dots, \sigma_N)$ , with the parameter  $J$  associated with an “external magnetic field” and  $E$  with the “nearest-neighbours interaction”,  $N$  the number of the atoms. The ground state is the lowest energy configuration  $s_g$  (in physics the zero temperature configuration), so that  $s_g \in \operatorname{argmin}_s H(s)$ . In a social network can be seen as the state with the most likely opinion, minimizing conflicts among its members (atoms).

In the standard bootstrap percolation process [16] a node is initially either active with a given probability  $f$  or inactive. It becomes active if  $k$  ( $k = 2, 3, \dots$ ) of its nearest neighbors are active. In that notion it resembles the  $k$ -core problem of random graphs [124], where  $k$ -core is the maximal subgraph within which all vertices have at least  $k$  neighbors, but whereas bootstrap percolation starts from a subset of seed vertices according to the above-mentioned activation rule, the  $k$ -core of the network can be found by a subsequent pruning of vertices which have less than  $k$  neighbors.

**Empirical Models.** Before the advent of machine-readable traces, the potential of networks in the transmission of information and messages was stated already by Milgram in his renowned experiment [130] or Christakis [79], who suggested in a study of 12000 participants that risks, such as the risk of becoming obese or benefits, such as stopping of smoking, are propagated through social ties. However, it is large scale and time-resolved machine-readable traces that, through the step-by-step track of interactions in OSNs (although not compulsorily easily accessible/collectible), have driven to the formulation of a plethora of empirical models.

Some generic observations concerning the empirical models are the following. Many of them lack insight of information content, unlike works, such as that of Huberman et al. [24], who formulate a model taking into consideration solely the features of an information item (a news item in Twitter). Sometimes the discovered patterns in empirical models are at odds with the predictions based on theoretical (analytical) models. For example, in unison with the epidemical model, Leskovec et al. in [119] claim that cascades (depicting the blogosphere information diffusion) are mostly tree-like. More specifically, they notice that the number of edges in the cascade increases almost linearly with the number of nodes, suggesting that the average degree in the cascade remains constant as the cascade grows (a trees property). Moreover, Leskovec et al. claim that these trees are balanced, as they notice that the cascade diameter increases logarithmically with the size of the cascade. In contradiction to the above, the trees derived from the chain-letter diffusion model of Liben-Nowell and Kleinberg in [121] are inconsistent with the epidemic model, as they are very narrow and deep, with the majority of their nodes having one child and a median distance from their root to the their leaves being of hundreds steps.

Precisely, in [121] the spread of a chain-letter is represented by a tree. Copies of the chain-letter represent paths through the tree, the root represents the originator and the leaves represent the recipients of a message ( $w$  is a child of  $v$  if  $w$  appends its name in the copy of the letter directly below  $v$ ). In order to produce trees with the characteristics mentioned in the previous paragraph, the probabilistic model suggested (i) incorporates asynchrony: after receiving a message, each recipient waits for a time  $t$  before acting on it, and if it receives more copies of the item in this time interval, it acts upon only one of them, and (ii) encompasses a back-rate  $\beta$ , as a node can either forward the message to its neighbours with probability  $1 - \beta$  or group-reply to his corecipients with a probability  $\beta$ .

In [27], Bakshy et al. attempt to model the information diffusion in Twitter with the use of regression trees. Twitter is convenient for information diffusion modeling, since it is explicitly diffusion-oriented: users subscribe to the content of other users. The retweet feature, moreover, helps in the acknowledgement (though does not guarantee it) of reposts. Seeders are users posting original (not retweeted) content and reposting instead of the conventional retweeting (RT @username) is taken into account. Influence is measured in terms of the size of the whole diffusion tree created, and not just the plain number of explicit retweets. The three different cases studied ascribe the influence to the first one having posted a link, the most recent one or follow a hybrid approach.

As far as the seed users are concerned, the predictors used include: the number of followers, number of friends, number of tweets and date of joining, and regarding the past influence of seed users: the average, minimum and maximum total influence and average, minimum and maximum local influence (local refers to the average number of reposts by a user’s immediate friends over a period of one month and total to the average total cascade size over that period).

Bakshy et al. [27] come to the conclusion that although large cascades have in their majority previously successful individuals with many followers as initiators, individuals with these characteristics are not necessarily bound to start a large cascade. Thus, because of the fact that estimations cannot be made at an individual level, marketers should rely on the average performance. By studying the return on investment, on the whole, with a cost function of the number of followers per individual  $i$ :  $c_i = ac_f + f_i c_f$ , where  $a$  is acquisition cost  $c_f$  cost per follower and  $f_i$  is the number of followers, they conclude that relatively ordinary users of average influence and connectivity are most cost-efficient.

Content-related features are, also, according to Bakshy et al. not expected to discriminate initiators of large cascades from non-successful ones, due to the large number of non-successes. In order to take content into account, the regression analysis is repeated encompassing the following features: rated interestingness, perceived interestingness to an average person, rated positive feeling, willingness to share via email, IM, Twitter, Facebook or Digg, some indicator variables for type of URL, and some indicator variables for category of content.

Moreover, Lerman et al. [118] claim that exploiting the proximity of users in the social graph can serve as an adding-value factor for the prediction of information diffusion. They discriminate proximity as coming from conservative or non-conservative processes (denoting that the amount of spread information in the network remains or not constant, respectively). For the case the underlying network is not fully known [132], Najjar et al. focus on predicting the final activation state of the network when an initial activation is given. They find the correspondence between the initial and final states of the network without considering the intermediate states. Their work is based on the analogy between predictive and generative approaches for discrimination or regression problems (predictive models depicting a better performance, when the real data distribution can’t be captured).

In [174], Yang and Leskovec use a time series model for modeling the global influence of a node through the whole network. For each node  $u$ , an influence function  $I_u(l)$  is the number of mentions of an information  $l$  time units after the node  $u$  adopted the information (at  $t_u$ ), and with  $V(t)$  being the number of nodes that mention the information at time  $t$ , it applies:

$$V(t + 1) = \sum_{u \in A(t)} I_u(t - t_u) \quad (5)$$

where  $A(t)$  are the nodes that got activated before  $t$ ,  $t_u \leq t$ . For the modeling of the influence functions a non-parametric formulation followed allows greater accuracy and deviation, as no assumptions are made.

A study of the social news aggregator Digg [66] crawling data from the site, story, user and social network perspective, suggests the presence of previously unconsidered factors for the steering of information spread in OSNs. Doerr et al. suggest, that, beyond the bare OSN topology two factors matter: the temporal alignment between user activities (i.e. whether users are visiting the site in the same narrow time window) and a hidden logical layer of interaction patterns occurring in their majority outside the social graph.

In the direction of studying the information diffusion as social graphs evolve, Ren et al. [146] study the evolution steps for shortest paths between two nodes, (so that they can ascribe them to a disjoint path, a short-circuiting bridge or a new friend between them), and furthermore, metrics such as closeness centrality, and global metrics, like the graph diameter, across snapshots of gradually evolving graphs. To this end, they adopt an efficient algorithm and an efficient storage scheme.

Firstly, they cluster (in an incremental procedure not requiring all snapshots to be present in memory) successive graphs exploiting their many resemblances (daily snapshots). As  $G_U$  and  $G_\cap$  essentially “bound” the graphs in the cluster, with  $G_\cap$  being the intersection (the largest common subgraph) of all snapshots in cluster  $C$ , and  $G_U$  the union (the smallest common supergraph) of all snapshots in  $C$ , grouping of snapshots into clusters can be based in the idea of the graph edit similarity between these two graphs ( $G_U, G_\cap$ ). The graph edit similarity to capture the similarity requirement of a cluster is defined as:

$$ges(G_a, G_b) = \frac{2 | E(G_a \cap G_b) |}{| E(G_a) | + | E(G_b) |} \quad (6)$$

Secondly, they exploit the idea that, denoting the shortest-path between the vertices  $v$  and  $u$ , by  $\tilde{P}_*(u, v)$  in a graph  $G_*$ , where  $* = 1, 2, \dots, n, \cap, \cup$ , the solution can easily be found by the intersection or union (two graphs) of graphs in the cluster, or be “fixed” using these two graphs, and they propose a “finding-verifying-fixing framework”.

As far as the storage schemes variations are concerned, for a cluster of snapshots  $C = G_1, \dots, G_k$  the deltas  $\Delta(G_i, G_\cap), \forall 1 \leq i \leq k$  consist a small fraction of the snapshot, and their size depends on the threshold value used for clusters’ similarity. The penalty of decompression overheads needed is surpassed by savings in I/O. Variations of the storage schemes include the following:

$$SM1(C) = \{G_\cap, \Delta(G_U, G_\cap), \Delta(G_i, G_\cap) | 1 \leq i \leq k\} \quad (7)$$

$$SM2(C) = \{G_\cap, \Delta(G_U, G_\cap), \Delta(G_1, G_\cap), \mathcal{D}(G_i, G_{i-1}) | 2 \leq i \leq k\} \quad (8)$$

$$SM\_FVF(C) = \{\mathcal{D}(G_\cap, G_{p\cap}), \Delta(G_U, G_\cap), \Delta(G_1, G_\cap), \mathcal{D}(G_i, G_{i-1}) | 2 \leq i \leq k\} \quad (9)$$

In (7) the authors consider only the edge sets of  $\Delta(G_i, G_\cap)$  and  $G_\cap$  to execute their algorithms on a snapshot  $G_i$  and the snapshots,  $G_i$ ’s, of the cluster need not be explicitly stored. For further compression of data of an evolving graph sequence similarity of successive snapshots is exploited: In (8)  $D(G_i, G_{i-1}) =$



$(E_i^+, E_i^-)$ , where  $E_i^+ = E(G_i) - E(G_{i-1})$  and  $E_i^- = E(G_{i-1}) - E(G_i)$  are the changes made to snapshot  $G_{i-1}$  to obtain the next snapshot  $G_i$ . Authors observe that the size of the set of edge changes  $D(G_i, G_{i-1})$  is on average just 1/10 the size of  $\Delta(G_i, G_{i-1})$ . Hence, representing an EGS in terms of the  $D$ 's is much more space efficient than in terms of the  $\Delta$ 's. Further compression can be achieved by exploiting inter-cluster redundancy (9).

## 6 Distribution of Social Data Streams

### 6.1 Content Distribution for Social Data Streams

This subsection provides a description of architectures, systems and techniques [108] for the distribution of social data content.

**Architectures.** In [99], Jacobson et al. introduce Content Centric Networking (CCN), noting that network use has evolved to be dominated by content distribution and retrieval. CCN has no notion of host at its lowest level - a packet "address" names content, not location, while simultaneously preserving the design decisions that make TCP/IP simple, robust and scalable. Content is treated as a primitive, and with new approaches, Jacobson et al. simultaneously achieve scalability and performance.

To share resources within the context of a social network with the use of the cloud business model, Chard et al. in [49] propose the SocialCloud architecture. Users register in cloud services (computational capacity, photo storage etc.), and their friends can consume and provide these services through a Facebook application. The allocation of resources (trading or reciprocal use between friends) is conducted by an underlying market infrastructure, whereas the Social Cloud application passes a SLA to the service. The advertisement of the service, so that it can be included in the market is done with XML based metadata stored in Globus Monitoring and Discovery System (MDS).

An interesting approach [116] applicable to the realm of content delivery is based on an architecture which combines global learning and local caches with small population. It is shown that age-based thresholds can timely exploit time-varying popularities to improve caching performance. Moreover, the caching efficiency is maximized by a combination of global learning and clustering of access locations, accompanied by score mechanisms to help with practical issues at local caches. Practical considerations include, though, the size of the content that circulates over OSN and the long-tail effect, since the goal of the authors is first to learn a good estimate at the global point and then feed it back to the local caches in the form of content scores, thus, making the approach possibly prohibitive for OSN-aware content delivery.

**Systems.** In Buzztraq [151], Sastry et al. build a prototype system that takes advantage of the knowledge of the users' friends' location and number, to generate hints for the placement of replicas closer to future accesses. Comparing their

strategy with location based placement, which instead uses the geographical location of recent users, they find substantial decrease of cost, when requests as part of cascades are more than random accesses of content. Furthermore, their system reacts faster when there is a new region shift, since it starts counting friends of previous users in a new region, even before a request comes from that region. The key concept of Buzztraq is to place replicas of items already posted by a user closer to the locations of friends, anticipating future requests. The intuition is that social cascades are rapidly spread through populations as social epidemics. The experimental results indicated that social cascade prediction can lower the cost of user access compared to simple location-based placement. Buzztrack is a simple system that only provides hints as to where to place objects. Other more complex constraints that the present work covers, such as server bandwidth and storage, are not taken into account. Moreover, social cascade is indirectly analyzed because there has to be a third-party page where users connect to view the videos and have access to their social profile.

In the direction of distributing long-tailed content while lowering bandwidth costs and improving QoS, although without considering storage constraints, Traverso et al. in [165] exploit the time differences between sites and the access patterns that users follow. Rather than naively pushing UGC immediately, which may not be consumed and contribute unnecessarily to a traffic spike in the upload link, the system can follow a pull-based approach, where the first friend of a user in a Point of Presence (PoP) asks for the content. Moreover, rather than pushing content as soon as a user uploads, content can be pushed at the local time that is off-peak for the uplink and be downloaded in a subsequent time bin, also off-peak for the downlink. The larger the difference is between the content production bin and the bin in which the content is likely to be read, the better is the performance of the system.

In [153], Scellato et al. study how Twitter can be used to examine social cascades of UGC from YouTube and discover popular objects for replication. They improve the temporary caching policy by placing content after accounting for the distance between users. For the model CDN system constructed and tested, Scellato et al. used the Limelight network properties with 19 clusters of servers worldwide. To test the system, two different video weights were used: geosocial, in which node locality values are calculated from all the users that have posted a message about the item (even without being involved in a cascade), and geocascade, in which node locality values are calculated from the users participating in the item's social cascade. It was shown that the model improved performance against a no weight policy, with geocascade weight performing better.

**Techniques.** The introduction of concrete, unified metrics for the characterization of the extent of the social dissemination (local or global cascades phenomena) is an open issue. A systematic incorporation of this quantified knowledge into the existent underlying content delivery infrastructure would be salutary for proactive steps towards the improvement of user experience.

Furthermore, novel techniques aim to incorporate the information extracted from OSNs in the way that users share content and in how the content ultimately reaches the users. Some of these works use the information directly from OSNs, whereas others use such information indirectly. The research goals vary: the decision for copying content, improvement of policy for temporary caching, etc.

Zhou et al. [178] leverage the connection between content exchange and geographic locality (using a Facebook dataset they identify significant geographic locality not only concerning the connections in the social graph, but also the exchange of content) and the observation that an important fraction of content is “created at the edge” (*is user-generated*), with a web based scheme for caching using the access patterns of friends. Content exchange is kept within the same Internet Service Provider (ISP) with a drop-in component, that can be deployed by existing web browsers and is independent of the type of content exchanged. Browsing users online are protected with  $k$ -anonymity, where  $k$  is the number of users connected to the same proxy and are able to view the content.

In [92], Hoque and Gupta propose a technique with a logical addressing scheme for putting together in the disk blocks containing data from friends. The large scale of OSNs and the predominant tail effect do not allow use of techniques such as those used in multimedia file systems or web servers, where items are globally popular, and, techniques keeping related blocks together tracking the access pattern of blocks, respectively. To this purpose, in [92] the social graph is divided into communities. The organization of blocks in the disk is conducted with a greedy heuristic that finds a layout for the users within the communities and organizes the different communities on the disk by considering inter-community tie strength. The system is implemented on top of the Neo4j graph database as a layout manager.

Instead of optimizing the performance of UGC services exploiting spatial and temporal locality in access patterns, Huguenin et al., in [95], show on a large (more than 650,000 videos) YouTube dataset that content locality (induced by the related videos feature) and geographic locality are in fact correlated. More specifically, they show how the geographic view distribution of a video can be inferred to a large extent from that of its related videos, proposing a UGC storage system that proactively places videos close to the expected requests. Such an approach could be extended with the leverage of information from OSNs.

Kilanioti et al. in [109–111] propose miscellaneous policies for dynamic OSN-aware content delivery over a content delivery simulation framework. The authors propose policies that take patterns of user activity over OSNs and exploit geo-social properties of users participating in social cascades, proceed to incorporate various caching schemes of the underlying infrastructure, different policies for the handling of OSN data and various approaches that take into account the efficient timing of prefetching. Given an efficient placement of surrogate servers with maximum performance and minimum infrastructure cost, they apply contextual features of the user as heuristics to find the best content diffusion placement, either in a global or in a local scale, i.e., which content will be copied in the surrogate servers and to what extent, not overlooking memory, time and

computational cost. Moreover they study temporal aspects of diffusion, related to the most efficient timing of the content placement. The simulation framework they introduce can serve as the basis of further parameterized content delivery experimentation that exploits information transmission over OSNs and decreases replication costs by selectively copying items to locations where items are likely to be consumed.

In terms of performance, Kilanioti et al. note a significant improvement over the respective improvement (39.43% only for the plain Social Prefetcher approach [107], up to 42.32% for selected caching mechanisms, compared to 30% in [165]) performing better than existent pull-based methods employed by most CDNs, even though these methods additionally overlook storage issues of the distributed infrastructure.

Last but not least, of more concurrent cascades happening it would be interesting to know which of them will evolve as global and which of them will evolve as local, possibly making some associations with their content or context features. It is challenging to discover contextual associations among the topics, which are by nature implicit in the user-generated content exchanged over OSNs and spread via social cascades. In other words it would be useful to derive semantic relations. This way the identification of a popular topic can be conducted in a higher, more abstract level with the augmentation of a semantic annotation. While the topic of a single information disseminated through an OSN can be explicitly identified, it is not trivial to identify reliable and effective models for the adoption of topics as time evolves [80,123] characterized with some useful emergent semantics. Therefore efficient semantic annotation can be seen as a solution for the challenge of characterization of the extent of the social dissemination.

## 6.2 Content Distribution in 5G Environments and Technologies

Content became the main information item exchanged between different actors in the Internet. Video and multimedia content counts for 80–90% of the total global traffic. Rich multimedia content lead to rapid mobile traffic growth that current mobile radio network, mobile backhaul, the capacity of the wireless link and mobile core network cannot support. 5G could overcome these bottlenecks introducing high increasing ratio of mobility communications and strong orientation towards content-related services and applications for content delivery over wireless technology, high throughput, low data delivery latency, and high scalability enabling huge number of devices [17].

**Environment.** 5G represents the 5th generation network of mobile systems which opens new possibilities, increase radio link capacity and brings plenty of new trends such as [17,122]: heterogeneous networks (HetNets); new use cases based on connections and communications between device to device, massive Machine-Type Communications, and Internet of Things (IoT); evolution of radio access technologies; cloudification throughout SDN and network function virtualization (NFV) paradigms; flexible spectrum management; cell densification; etc.

NFV and SDN capabilities in 5G systems are expected to enable network programmability. Content delivery could be affected by 5G cloudification through different SDN/NFV paradigms [122]. Programmable network control and the virtualization of all the RAN elements into virtual appliances by flexible NFV management are included within 5G networks. This enable content focused resources allocation. Agile design of new network functions and their control are possible. Network providers could extend network with new function that includes custom designed information, such as services that can offer to the online media service providers. The collaboration between the network provider and the online media service provider by means of the edge cache could be enabled by media delivery solutions designed for 5G. The control of the network will be kept by the network provider. The network provider would give only the relevant information for the online media service provider, while the online media service provider will keep the control of the delivery process and decide whether the cache shall be used, what and how information or resources are cached [3].

**Technologies.** New technologies such as LTE-A, LTE-U, WiFi and ZigBee, SDN and NFV rapidly change networks and services and lead to changes to content delivery. For example, mobile video will generate more than 69% of mobile data traffic by 2019 [2]. It is expected to witness an increase to 75% by 2021, which is much greater from 46% in 2016 on the share of smart devices and connections, while the amount of traffic offloaded from 4G was 63% at the end of 2016, and it will be 66% percent by 2021 [2]. It is also expected to witness higher offload rates when the 5G network arrives. The main challenges in a wireless or mobile environment that have impact on content delivery services are reflected to the limited spectrum and bandwidth in wireless, heterogeneous networks, wireless link characteristics that are dependent on location and time, radio congestion, handoff issues, etc.

**5G Use Cases.** Future 5G developments is dependent on service providers, technology enablers and customers. All these actors are directly involved in decisions which use cases to pursue first, as well what technology is needed for the use cases. The 5G standards development process is also dependent on decisions which use cases first to deploy. All these ongoing developments will directly affect content delivery mechanisms, models and systems architectures. The main 5G use cases currently are reflected to [3]:

1. Gigabit broadband to home, related to deliver streams rated from 100 Mbit/sec to 1 Gbit/sec, which are needed to deliver television with higher resolution than 4K, virtual and augmented reality. Specific applications require special network configuration, for example in order to minimize latency in virtual reality applications.
2. Next generation mobile user experience.
3. Future corporate networks, addressed to better service providing, which require operators to dynamically manage network and to use software defined networking and network function virtualization.

4. Digital industrial ecosystems include agriculture, smart cities and healthcare applications, which imply network configurations that every industry participant can benefit from.
5. Infrastructure as a service approach is for service providers that lack the resources to invest in nationwide 5G coverage.

**5G Solutions and Approaches.** The demand for high-speed data applications has been on the rise in the recent decade, which led to the development of 5G. Development of efficient mechanisms for supporting mobile multimedia and data services is prerequisite for 5G networks. The real bottleneck of today's mobile networks is access radio network and the backhaul. Caching in the intermediate nodes, servers, gateways, routers, and mobile users' devices can reduce doubled transmission from content providers and core mobile networks.

Known caching techniques that can be used within 5G are: content distribution network, information-centric networks, content-centric networking, http web caching, evolved packet core caching, radio access network caching, device to device caching, proactive caching, predictive caching, cooperative caching [23]. Those techniques are using different algorithms and models. Analysis presented in [23] showed that the deployment of those caching techniques in mobile network can reduce redundant traffic in backhaul, minimize the traffic load, increase the transfer rate in mobile network and reduce the latency. Correlation of several caching methods and procedures could result in improving network performance and obtaining better results.

5G brings complex heterogeneity of the network with different technologies that coexist, where some technologies could totally disable transmission of data of equipment that use other technologies. Solutions that efficiently handles resources in space, frequency, and device dimensions are needed. One possible efficient solution is semantic coordination in such networks is given in [135, 164].

The nodes in the system can communicate and share knowledge of their perspective of the spectrum utilization in the network. In [164] authors proposed to model the spectrum usage coordination as an interactive process between a number of distributed communicating agents, where agents share their specific information and knowledge. The information includes the current spectrum usage state, spatial coordinates of the device, available communication protocols, usage policy, spectrum sensing capabilities of the device, spectrum needs, etc. Approach for such coordination presented in [164] is based on semantic technologies, and harmonize communication between heterogeneous agents with potentially different capabilities with a minimal common compliance. The core knowledge is represented by ontologies whose representation and usage is specified in a standardized way. This semantic technologies-based approach can be used for a wide spectrum of problems within 5G heterogeneous networks, such as network states predictions, network analysis, minimizing traffic load, content distribution coordination etc. This approach could be used in combination with caching techniques in order to improve content distribution in 5G, but further research should be carried out in this area.

## 7 Conclusions

This article describes the results of the collaborative work performed as part of High-Performance Modelling and Simulation for Big Data Applications (cHiPSet) COST Action 1406. The presented case study focused on multimedia big data from entertainment and social media, medical images, consumer images, voice and video, that drives research and development of related technologies and applications and is steadily becoming a valuable source of information and insights [109–111], [108]. In fact, this work describes the general landscape and how our approach fits in the general ecosystem. Multimedia content providers such as YouTube strive to efficiently deliver multimedia big data to a large amount of users over the Internet, with currently more than 300 h of video content being uploaded to the site every minute. Traditionally, these content providers often rely on social data content distribution infrastructures. However, some measurement studies depict that a significantly large proportion of HTTP traffic results from bandwidth-intensive multimedia content circulating through OSNs. Consequently the user activity extracted from OSNs can be exploited to reduce the bandwidth usage. By incorporating patterns of information transmission over OSNs into a simulated content distribution infrastructure, the performance of content distribution mechanisms can be remarkably improved.

CDN services are increasingly being used to enable the delivery of bandwidth-demanding large media data to end-users of multimedia content providers and extend the capabilities of the Internet by deploying massively distributed infrastructures to accelerate content delivery. Next generation CDNs are being leveraged in an array of ways to overcome the challenges of providing a seamless customer experience across multiple devices with varying connectivity and corresponding to the call for enterprise application delivery. They have to go beyond efficient resource discovery and retrieval tasks of the established CDNs and support refined mechanisms for data placement, replication and distribution for a large variety of resource types and media formats. OSNs on the other hand create a potentially transformational change in user navigation and from this angle the rapid proliferation of OSNs sites is expected to reshape the architecture and design of CDNs. The challenges and opportunities highlighted in the interdisciplinary field of OSN-aware content delivery are bound to foster some interesting future developments, including innovative cache replacement strategies as a product of the systematic research of temporal, structural and geographical properties of social cascades.

Particularly today that HTTP traffic ascribed to media circulating over OSNs has grown, an OSN-awareness mechanism over content distribution schemes has become essential. This mechanism aims to exploit patterns of social interactions of the users to reduce the load on the origin server, the traffic on the Internet, and ultimately improve the user experience. By addressing the issue of which content will be copied in the surrogate servers of a CDN, it ensures a near-optimal content diffusion placement. At the same time, it moderates the impact on bandwidth that the Big Data transmitted via OSNs has, offering scalable

solutions to existing CDNs or OSNs providers. Furthermore, it paves the way for experimentation with variations on caching schemes, timing parameters of content delivery and context of the OSN and the media platform.

A future target is to potentially leverage the CDN services of cloud service providers in order to lower costs while increasing simplicity. CDNs, often operated as Software as a Service (SaaS) in cloud providers (Amazon CloudFront, Microsoft Azure CDN, etc.) aim at addressing the problem of smooth and transparent content delivery. A CDN actually drives cloud adoption through enhanced performance, scalability and cost reduction. With the limitation for both CDNs and cloud services being the geographic distance between a user asking for content and the server where the content resides, cloud acceleration and CDN networks are both complementary to achieving a goal of delivering data in the fastest possible way. Cloud mainly handles. Utilization of OSN-aware CDNs in cloud computing, where content is constantly changing and, thus, not easily cached, is likely to have profound effects on large data download.

## References

1. Alexa. <http://alexa.com/topsites>. Accessed 20 Dec 2018
2. Cisco visual networking index: global mobile data traffic forecast update, 2016–2021 white paper
3. Ericsson research blog. 5G media delivery. <https://www.ericsson.com/research-blog/5g-media-delivery/>. Accessed 20 Dec 2018
4. Facebook Newsroom. <http://newsroom.fb.com/Key-Facts>. Accessed 20 Dec 2018
5. International Telecommunication Union. ICT facts and figures in 2017. <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf>. Accessed 20 Dec 2018
6. Internet Society. Global Internet Report 2017. Mobile Evolution and Development of the Internet. <https://future.internetsociety.org/wp-content/uploads/2017/09/2017-Internet-Society-Global-Internet-Report-Paths-to-Our-Digital-Future.pdf>. Accessed 20 Dec 2018
7. Management of Networks with Constrained Devices: Use Cases. IETF Internet Draft. <https://goo.gl/cT5pXr>. Accessed 15 Mar 2016
8. Mobile-edge computing - introductory technical white paper. <https://goo.gl/ybrCnq>. Accessed 15 Mar 2016
9. The Internet of Things: how the next evolution of the Internet is changing everything. CISCO, San Jose, CA, USA, White Paper (2011). <http://goo.gl/ugKAoN>. Accessed 15 Mar 2016
10. The Smart Grid: An Introduction. US Department of Energy. <http://goo.gl/jTNgf>. Accessed 15 Mar 2016
11. Under the Hood: Scheduling MapReduce jobs more efficiently with Corona, Facebook engineering. [goo.gl/XJRNN](http://goo.gl/XJRNN). Accessed 20 Dec 2018
12. YouTube Statistics. <https://www.youtube.com/yt/press/statistics.html>. Accessed 20 Dec 2018
13. cHiPSet - Research Work Results. Grant Period 1 (2017). <http://chipset-cost.eu/wp-content/uploads/2017/05/report-1.pdf#page=13>. Accessed 20 Dec 2018
14. Abedini, N., Shakkottai, S.: Content caching and scheduling in wireless networks with elastic and inelastic traffic. *IEEE/ACM Trans. Netw.* **22**(3), 864–874 (2014)



15. Abolfazli, S., Sanaei, Z., Ahmed, E., Gani, A., Buyya, R.: Cloud-based augmentation for mobile devices: motivation, taxonomies, and open challenges. *IEEE Commun. Surv. Tutor.* **16**(1), 337–368 (2014). <https://doi.org/10.1109/SURV.2013.070813.00285>
16. Aizenman, M., Lebowitz, J.L.: Metastability effects in bootstrap percolation. *J. Phys. A: Math. Gen.* **21**(19), 3801 (1999)
17. Akpakwu, G.A., Silva, B.J., Hancke, G.P., Abu-Mahfouz, A.M.: A survey on 5G networks for the internet of things: communication technologies and challenges. *IEEE Access* **6**, 3619–3647 (2018)
18. Amur, H., Cipar, J., Gupta, V., Ganger, G.R., Kozuch, M.A., Schwan, K.: Robust and flexible power-proportional storage. In: *Proceedings of the 1st ACM Symposium on Cloud Computing*, pp. 217–228. ACM (2010)
19. Ananthanarayanan, G., et al.: PACMan: coordinated memory caching for parallel jobs. In: *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, p. 20. USENIX Association (2012)
20. Anderson, C.: *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion, Santa Clara (2008). Revised and Updated Edition
21. Anjum, N., Karamshuk, D., Shikh-Bahaei, M., Sastry, N.: Survey on peer-assisted content delivery networks. *Comput. Netw.* **116**, 79–95 (2017)
22. Arthur, W.B.: Competing technologies, increasing returns, and lock-in by historical events. *Econ. J.* **99**(394), 116–131 (1989)
23. Assila, B., Kobbane, A., El Koutbi, M.: A survey on caching in 5G mobile network (2017)
24. Asur, R.B.S., Bandari, R., Huberman, B.: The pulse of news in social media: forecasting popularity. In: *Association for the Advancement of Artificial Intelligence*, 1202 (2012)
25. Avelar, V., Azevedo, D., French, A., Power, E.N.: PUE: a comprehensive examination of the metric. *White Paper* 49 (2012)
26. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone’s an influencer: quantifying influence on Twitter. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 65–74. ACM (2011)
27. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone’s an influencer: quantifying influence on Twitter. In: *Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, 9–12 February 2011*, pp. 65–74 (2011). <http://doi.acm.org/10.1145/1935826.1935845>
28. Bakshy, E., Rosenn, I., Marlow, C., Adamic, L.A.: The role of social networks in information diffusion. In: *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, 16–20 April 2012*, pp. 519–528 (2012). <http://doi.acm.org/10.1145/2187836.2187907>
29. Banerjee, A.V.: A simple model of herd behavior. *Q. J. Econ.* **107**(3), 797–817 (1992)
30. Barroso, L.A., Clidaras, J., Hölzle, U.: The datacenter as a computer: an introduction to the design of warehouse-scale machines. *Synth. Lect. Comput. Arch.* **8**(3), 1–154 (2013)
31. Bashroush, R.: A comprehensive reasoning framework for hardware refresh in data centers. *IEEE Trans. Sustain. Comput.* **3**(4), 209–220 (2018)
32. Bashroush, R., Woods, E.: Architectural principles for energy-aware internet-scale applications. *IEEE Softw.* **34**(3), 14–17 (2017)
33. Bass, F.M.: A new product growth for model consumer durables. *Manag. Sci.* **15**(5), 215–227 (1969)

34. Beloglazov, A., Abawajy, J., Buyya, R.: Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Futur. Gener. Comput. Syst.* **28**(5), 755–768 (2012)
35. Beloglazov, A., Buyya, R.: Energy efficient resource management in virtualized cloud data centers. In: *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, pp. 826–831. IEEE Computer Society (2010)
36. Beloglazov, A., Buyya, R.: Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. *Concurr. Comput.: Pract. Exp.* **24**(13), 1397–1420 (2012)
37. Berger, E.: Dynamic monopolies of constant size. *J. Comb. Theory Ser. B* **83**(2), 191–200 (2001)
38. Bhattacharya, A.A., Culler, D., Friedman, E., Ghodsi, A., Shenker, S., Stoica, I.: Hierarchical scheduling for diverse datacenter workloads. In: *Proceedings of the 4th Annual Symposium on Cloud Computing*, p. 4. ACM (2013)
39. Bikhchandani, S., Hirshleifer, D., Welch, I.: A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100**, 992–1026 (1992)
40. Bonacich, P.: Power and centrality: a family of measures. *Am. J. Sociol.* **92**, 1170–1182 (1987)
41. Borgatti, S.P.: Centrality and network flow. *Soc. Netw.* **27**(1), 55–71 (2005). <http://www.sciencedirect.com/science/article/pii/S0378873304000693>
42. Borthakur, D., et al.: Apache Hadoop goes realtime at Facebook. In: *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, pp. 1071–1080. ACM (2011)
43. Bottazzi, D., Montanari, R., Toninelli, A.: Context-aware middleware for anytime, anywhere social networks. *IEEE Intell. Syst.* **22**(5), 23–32 (2007)
44. Boyd, D., Ellison, N.B.: Social network sites: definition, history, and scholarship. *J. Comput.-Mediat. Commun.* **13**(1), 210–230 (2007). <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
45. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **30**(1), 107–117 (1998)
46. Brodersen, A., Scellato, S., Wattenhofer, M.: YouTube around the world: geographic popularity of videos. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 241–250. ACM (2012)
47. Van den Bulte, C., Joshi, Y.V.: New product diffusion with influentials and imitators. *Mark. Sci.* **26**(3), 400–421 (2007)
48. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., Moon, S.B.: I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system. In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement 2007, San Diego, California, USA, 24–26 October 2007*, pp. 1–14 (2007). <https://doi.org/10.1145/1298306.1298309>
49. Chard, K., Caton, S., Rana, O., Bubendorfer, K.: Social cloud: cloud computing in social networks. In: *IEEE International Conference on Cloud Computing, CLOUD 2010, Miami, FL, USA, 5–10 July 2010*, pp. 99–106 (2010). <https://doi.org/10.1109/CLOUD.2010.28>
50. Chen, Y., Alspaugh, S., Katz, R.: Interactive analytical processing in big data systems: a cross-industry study of mapreduce workloads. *Proc. VLDB Endow.* **5**(12), 1802–1813 (2012)
51. Chib, S., Greenberg, E.: Understanding the metropolis-hastings algorithm. *Am. Stat.* **49**(4), 327–335 (1995)

52. Clifford, P., Sudbury, A.: A model for spatial conflict. *Biometrika* **60**(3), 581–588 (1973)
53. Coutand, O., et al.: Context-aware group management in mobile environments. In: *IST Mobile Summit* (2005)
54. Daley, D., Kendall, D.G.: Stochastic rumours. *IMA J. Appl. Math.* **1**(1), 42–55 (1965)
55. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Commun. ACM* **51**(1), 107–113 (2008)
56. Delgado, P., Didona, D., Dinu, F., Zwaenepoel, W.: Job-aware scheduling in eagle: divide and stick to your probes. In: *Proceedings of the Seventh ACM Symposium on Cloud Computing*. No. EPFL-CONF-221125 (2016)
57. Delgado, P., Dinu, F., Kermarrec, A.M., Zwaenepoel, W.: Hawk: hybrid datacenter scheduling. In: *USENIX Annual Technical Conference*, pp. 499–510 (2015)
58. Delimitrou, C., Kozyrakis, C.: Paragon: QoS-aware scheduling for heterogeneous datacenters. *ACM SIGPLAN Not.* **48**, 77–88 (2013)
59. Delimitrou, C., Kozyrakis, C.: Quasar: resource-efficient and QoS-aware cluster management. *ACM SIGPLAN Not.* **49**, 127–144 (2014)
60. Delimitrou, C., Sanchez, D., Kozyrakis, C.: Tarcil: reconciling scheduling speed and quality in large shared clusters. In: *Proceedings of the Sixth ACM Symposium on Cloud Computing*, pp. 97–110. ACM (2015)
61. Dewi, R.R., Kim, T.-H.: A new approach to modeling of bio-inspired information diffusion with ant colony optimization in complex networks. In: Zelinka, I., Suganthan, P.N., Chen, G., Snasel, V., Abraham, A., Rössler, O. (eds.) *Nostradamus 2014: Prediction, Modeling and Analysis of Complex Systems*. AISC, vol. 289, pp. 131–140. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-07401-6\\_13](https://doi.org/10.1007/978-3-319-07401-6_13)
62. Dey, A.K.: Understanding and using context. *Pers. Ubiquit. Comput.* **5**(1), 4–7 (2001)
63. Dey, A.K., Abowd, G.D., Salber, D.: A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Hum.-Comput. Interact.* **16**(2), 97–166 (2001)
64. Dickens, L., Molloy, I., Lobo, J., Cheng, P.C., Russo, A.: Learning stochastic models of information flow. In: *2012 IEEE 28th International Conference on Data Engineering (ICDE)*, pp. 570–581. IEEE (2012)
65. Dodge, Y., Cox, D., Commenges, D., Davison, A., Solomon, P., Wilson, S.: *The Oxford Dictionary of Statistical Terms*. Oxford University Press, Oxford (2006)
66. Doerr, C., Blenn, N., Tang, S., Van Mieghem, P.: Are friends overrated? A study for the social news aggregator digg.com. *Comput. Commun.* **35**(7), 796–809 (2012)
67. Dogar, F.R., Karagiannis, T., Ballani, H., Rowstron, A.: Decentralized task-aware scheduling for data center networks. In: *ACM SIGCOMM Computer Communication Review*, vol. 44, pp. 431–442. ACM (2014)
68. Duy, T.V.T., Sato, Y., Inoguchi, Y.: Performance evaluation of a green scheduling algorithm for energy savings in cloud computing. In: *2010 IEEE International Symposium on Parallel & Distributed Processing, Workshops and PhD Forum (IPDPSW)*, pp. 1–8. IEEE (2010)
69. Easley, D.A., Kleinberg, J.M.: *Networks, Crowds, and Markets - Reasoning About a Highly Connected World*. Cambridge University Press (2010). [http://www.cambridge.org/gb/knowledge/isbn/item2705443/?site\\_locale=en\\_GB](http://www.cambridge.org/gb/knowledge/isbn/item2705443/?site_locale=en_GB)
70. Even-Dar, E., Shapira, A.: A note on maximizing the spread of influence in social networks. *Inf. Process. Lett.* **111**(4), 184–187 (2011). <https://doi.org/10.1016/j.ipl.2010.11.015>

71. Fard, A., Abdolrashidi, A., Ramaswamy, L., Miller, J.A.: Towards efficient query processing on massive time-evolving graphs. In: 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom 2012, Pittsburgh, PA, USA, 14–17 October 2012, pp. 567–574 (2012). <http://dx.doi.org/10.4108/icst.collaboratecom.2012.250532>
72. Fernández-Cerero, D., Jakóbič, A., Fernández-Montes, A., Kolodziej, J.: Stackelberg game-based models in energy-aware cloud scheduling. In: Nolle, L. (ed.) Proceedings of 32nd European Conference on Modelling and Simulation ECMS 2018 (ECMS, Wilhelmshaven, Germany, May 2018), ECMS 2018, pp. 460–467. European Council for Modelling and Simulation, Dudweiler (2018)
73. Fernández-Cerero, D., Jakóbič, A., Fernández-Montes, A., Kolodziej, J.: Game-score: game-based energy-aware cloud scheduler and simulator for computational clouds. *Simul. Model. Pract. Theory* (2018). <http://www.sciencedirect.com/science/article/pii/S1569190X18301229>
74. Fernández-Cerero, D., Fernández-Montes, A., Jakóbič, A., Kolodziej, J., Toro, M.: Score: simulator for cloud optimization of resources and energy consumption. *Simul. Model. Pract. Theory* **82**, 160–173 (2018). <http://www.sciencedirect.com/science/article/pii/S1569190X18300030>
75. Fernández-Cerero, D., Fernández-Montes, A., Ortega, J.A.: Energy policies for data-center monolithic schedulers. *Expert. Syst. Appl.* **110**, 170–181 (2018). <http://www.sciencedirect.com/science/article/pii/S0957417418303531>
76. Fernández-Cerero, D., Fernández-Montes, A., Velasco, F.: Productive efficiency of energy-aware data centers. *Energies* **11**(8), 2053 (2018). <http://www.mdpi.com/1996-1073/11/8/2053>
77. Fernández-Cerero, D., Jakóbič, A., Grzonka, D., Kolodziej, J., Fernández-Montes, A.: Security supportive energy-aware scheduling and energy policies for cloud environments. *J. Parallel Distrib. Comput.* **119**, 191–202 (2018). <http://www.sciencedirect.com/science/article/pii/S0743731518302843>
78. Fernández-Montes, A., Fernández-Cerero, D., González-Abril, L., Álvarez García, J.A., Ortega, J.A.: Energy wasting at internet data centers due to fear. *Pattern Recognit. Lett.* **67**, 59–65 (2015). <http://www.sciencedirect.com/science/article/pii/S0167865515001907>. *Cognitive Systems for Knowledge Discovery*
79. Fowler, J., Christakis, N.: *Connected: The Surprising Power of Our Social Networks and How they Shape Our Lives*. HarperCollins Publishers, New York City (2009)
80. García-Silva, A., Kang, J.H., Lerman, K., Corcho, O.: Characterising emergent semantics in Twitter lists. In: Proceedings of the 9th International Conference on the Semantic Web: Research and Applications (ESWC), Heraklion, Greece (2012)
81. Gea, T., Paradells, J., Lamarca, M., Roldan, D.: Smart cities as an application of Internet of Things: experiences and lessons learnt in Barcelona. In: 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), pp. 552–557. IEEE (2013)
82. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Mark. Lett.* **12**(3), 211–223 (2001)
83. Govindan, S., Liu, J., Kansal, A., Sivasubramanian, A.: Cuanta: quantifying effects of shared on-chip resource interference for consolidated virtual machines. In: Proceedings of the 2nd ACM Symposium on Cloud Computing, p. 22. ACM (2011)
84. Goyal, A., Bonchi, F., Lakshmanan, L.V.: Learning influence probabilities in social networks. In: Proceedings of the third ACM International Conference on Web Search and Data Mining, pp. 241–250. ACM (2010)

85. Grandl, R., Ananthanarayanan, G., Kandula, S., Rao, S., Akella, A.: Multi-resource packing for cluster schedulers. *ACM SIGCOMM Comput. Commun. Rev.* **44**(4), 455–466 (2015)
86. Granovetter, M.: Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978)
87. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through Blogspace. In: *Proceedings of the 13th International Conference on World Wide Web*, pp. 491–501. ACM (2004)
88. Henriksen, K., Indulska, J., Rakotonirainy, A.: Infrastructure for pervasive computing: challenges. In: *GI Jahrestagung (1)*, pp. 214–222 (2001)
89. Hindman, B., et al.: Mesos: a platform for fine-grained resource sharing in the data center. *NSDI* **11**, 22 (2011)
90. Hinze, A., Buchanan, G.: Context-awareness in mobile tourist information systems: challenges for user interaction. In: *International Workshop on Context in Mobile HCI at the Seventh International Conference on Human Computer Interaction with Mobile Devices and Services* (2005)
91. Holley, R.A., Liggett, T.M.: Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.* **3**, 643–663 (1975)
92. Hoque, I., Gupta, I.: Disk layout techniques for online social network data. *IEEE Internet Comput.* **16**(3), 24–36 (2012)
93. Hosseini-Pozveh, M., Zamanifar, K., Naghsh-Nilchi, A.R.: Assessing information diffusion models for influence maximization in signed social networks. *Expert Syst. Appl.* **119**, 476–490 (2019)
94. Hofffeld, T., Schatz, R., Biersack, E., Plissonneau, L.: Internet video delivery in YouTube: from traffic measurements to quality of experience. In: Biersack, E., Callegari, C., Matijasevic, M. (eds.) *Data Traffic Monitoring and Analysis*. LNCS, vol. 7754, pp. 264–301. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-36784-7\\_11](https://doi.org/10.1007/978-3-642-36784-7_11)
95. Huguenin, K., Kermarrec, A.M., Kloudas, K., Taïani, F.: Content and geographical locality in user-generated content sharing systems. In: *Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video*, pp. 77–82. ACM (2012)
96. Isard, M., Prabhakaran, V., Currey, J., Wieder, U., Talwar, K., Goldberg, A.: Quincy: fair scheduling for distributed computing clusters. In: *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*, pp. 261–276. ACM (2009)
97. Ising, E.: Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei* **31**(1), 253–258 (1925)
98. Istepanian, R., Hu, S., Philip, N., Sungoor, A.: The potential of Internet of m-health Things m-LoT for non-invasive glucose level sensing. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pp. 5264–5266. IEEE (2011)
99. Jacobson, V., Smetters, D.K., Thornton, J.D., Plass, M.F., Briggs, N., Braynard, R.: Networking named content. *Commun. ACM* **55**(1), 117–124 (2012). <https://doi.org/10.1145/2063176.2063204>
100. Jakóbi, A., Grzonka, D., Kołodziej, J.: Security supportive energy aware scheduling and scaling for cloud environments (2017)
101. Jiang, L., Feng, G., Qin, S.: Content distribution for 5G systems based on distributed cloud service network architecture (2015)

102. Juarez, F., Ejarque, J., Badia, R.M.: Dynamic energy-aware scheduling for parallel task-based application in cloud computing. *Futur. Gener. Comput. Syst.* **78**, 257–271 (2016)
103. Karanasos, K., et al.: Mercury: hybrid centralized and distributed scheduling in large shared clusters. In: *USENIX Annual Technical Conference*, pp. 485–497 (2015)
104. Katz, E., Lazarsfeld, P.F.: *Personal Influence, the Part Played by People in the Flow of Mass Communications*. Transaction Publishers, Piscataway (1966)
105. Kaushik, R.T., Bhandarkar, M.: GreenHDFS: towards an energy-conserving, storage-efficient, hybrid Hadoop compute cluster. In: *Proceedings of the USENIX Annual Technical Conference*, p. 109 (2010)
106. Kempe, D., Kleinberg, J.M., Tardos, É.: Maximizing the spread of influence through a social network. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146. ACM (2003)
107. Kilanioti, I.: Improving multimedia content delivery via augmentation with social information. The social prefetcher approach. *IEEE Trans. Multimed.* **17**(9), 1460–1470 (2015). <https://doi.org/10.1109/TMM.2015.2459658>
108. Kilanioti, I., et al.: A survey on cost-effective context-aware distribution of social data streams over energy-efficient data centres. *Simul. Model. Pract. Theory* (2018). <http://www.sciencedirect.com/science/article/pii/S1569190X18301709>
109. Kilanioti, I., Papadopoulos, G.A.: Delivering social multimedia content with scalability. In: Pop, F., Kołodziej, J., Di Martino, B. (eds.) *Resource Management for Big Data Platforms*. CCN, pp. 383–399. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-44881-7\\_18](https://doi.org/10.1007/978-3-319-44881-7_18)
110. Kilanioti, I., Papadopoulos, G.A.: Predicting video virality on Twitter. In: Pop, F., Kołodziej, J., Di Martino, B. (eds.) *Resource Management for Big Data Platforms*. CCN, pp. 419–439. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-44881-7\\_20](https://doi.org/10.1007/978-3-319-44881-7_20)
111. Kilanioti, I., Papadopoulos, G.A.: Content delivery simulations supported by social network-awareness. *Simul. Model. Pract. Theory* **71**, 114–133 (2017)
112. Kleinberg, J.M.: Cascading behavior in networks: algorithmic and economic issues. In: Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V. (eds.) *Algorithmic Game Theory*, pp. 613–632. Cambridge University Press (2007)
113. Ko, Y.Y., Cho, K.J., Kim, S.W.: Efficient and effective influence maximization in social networks: a hybrid-approach. *Inf. Sci.* **465**, 144–161 (2018)
114. Kuperman, M., Abramson, G.: Small world effect in an epidemiological model. *Phys. Rev. Lett.* **86**(13), 2909–2912 (2001)
115. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. *J. Supercomput.* **48**(9), 140–150 (2010). <https://doi.org/10.1109/MCOM.2010.5560598>
116. Leconte, M., Paschos, G., Gkatzikis, L., Draief, M., Vassilaras, S., Chouvardas, S.: Placing dynamic content in caches with small population. In: *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, April 2016
117. Lee, Y.C., Zomaya, A.Y.: Energy efficient utilization of resources in cloud computing systems. *J. Supercomput.* **60**(2), 268–280 (2012)
118. Lerman, K., Intagorn, S., Kang, J.H., Ghosh, R.: Using proximity to predict activity in social networks. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 555–556. ACM (2012)

119. Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M.: Patterns of cascading behavior in large blog graphs. In: Proceedings of SIAM International Conference on Data Mining (SDM) 2007. SIAM (2007)
120. Li, M., Wang, X., Gao, K., Zhang, S.: A survey on information diffusion in online social networks: models and methods. *Information* **8**(4), 118 (2017)
121. Liben-Nowell, D., Kleinberg, J.: Tracing information flow on a global scale using Internet chain-letter data. *Proc. Natl. Acad. Sci.* **105**(12), 4633–4638 (2008)
122. Liberal, F., Kourtis, A., Fajardo, J.O., Koumaras, H.: Multimedia content delivery in SDN and NFV-based towards 5G networks. *IEEE COMSOC MMTC E-Lett.* **10**(4), 6–10 (2015)
123. Lin, C.X., Mei, Q., Jiang, Y., Han, J., Qi, S.: Inferring the diffusion and evolution of topics in social communities. *Mind* **3**(d4), d5 (2011)
124. Luczak, T.: Size and connectivity of the K-core of a random graph. *Discrete Math.* **91**(1), 61–68 (1991)
125. Luo, X., Wang, Y., Zhang, Z., Wang, H.: Superset: a non-uniform replica placement strategy towards high-performance and cost-effective distributed storage service. In: 2013 International Conference on Advanced Cloud and Big Data (CBD), pp. 139–146. IEEE (2013)
126. Luu, M.D., Hoang, T.A., Lim, E.P.: A survey of information diffusion models and relevant problems (2011)
127. Maier, G., Feldmann, A., Paxson, V., Allman, M.: On dominant characteristics of residential broadband internet traffic. In: Proceedings of the 9th ACM SIGCOMM Internet Measurement Conference IMC, pp. 90–102. ACM (2009)
128. Mars, J., Tang, L.: Whare-map: heterogeneity in homogeneous warehouse-scale computers. In: ACM SIGARCH Computer Architecture News, vol. 41, pp. 619–630. ACM (2013)
129. Menon, A.: Big data@ Facebook. In: Proceedings of the 2012 Workshop on Management of Big Data Systems, pp. 31–32. ACM (2012)
130. Milgram, S.: The small world problem. *Psychol. Today* **2**(1), 60–67 (1967)
131. Morris, S.: Contagion. *Rev. Econ. Stud.* **67**(1), 57–78 (2000)
132. Najar, A., Denoyer, L., Gallinari, P.: Predicting information diffusion on social networks with partial knowledge. In: Proceedings of the 21st International Conference on World Wide Web, pp. 1197–1204. ACM (2012)
133. Nathuji, R., Kansal, A., Ghaffarkhah, A.: Q-clouds: managing performance interference effects for QoS-aware clouds. In: Proceedings of the 5th European Conference on Computer Systems, pp. 237–250. ACM (2010)
134. Navarro, N.D.A.B., Da Costa, C.A., Barbosa, J.L.V., Righi, R.D.R.: A context-aware spontaneous mobile social network. In: Ubiquitous Intelligence and Computing and 2015 IEEE 12th International Conference on Autonomic and Trusted Computing and 2015 IEEE 15th International Conference on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), pp. 85–92. IEEE (2015)
135. Nejković, V., Jelenković, F., Tošić, M., Milošević, N., Nikolić, Z.: CoordSS: an ontology framework for heterogeneous networks experimentation. *Telfor J.* **8**(2), 70–75 (2016)
136. Nekovee, M., Moreno, Y., Bianconi, G., Marsili, M.: Theory of rumour spreading in complex social networks. *Phys. A: Stat. Mech. its Appl.* **374**(1), 457–470 (2007)
137. Ousterhout, K., Wendell, P., Zaharia, M., Stoica, I.: Sparrow: distributed, low latency scheduling. In: Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, pp. 69–84. ACM (2013)

138. Pejovic, V., Musolesi, M.: Anticipatory mobile computing: a survey of the state of the art and research challenges. *ACM Comput. Surv.* **47**(3), 47 (2015). <https://doi.org/10.1145/2693843>
139. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Context aware computing for the internet of things: a survey. *IEEE Commun. Surv. Tutor.* **16**(1), 414–454 (2014)
140. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Sensing as a service model for smart cities supported by internet of things. *Trans. Emerg. Telecommun. Technol.* **25**(1), 81–93 (2014)
141. Plissonneau, L., Vu-Brugier, G.: Mobile data traffic analysis: how do you prefer watching videos? In: *Proceedings of the 22nd International Teletraffic Congress (ITC)*, pp. 1–8. IEEE (2010)
142. Qu, H., Mashayekhi, O., Terei, D., Levis, P.: Canary: a scheduling architecture for high performance cloud computing. arXiv preprint [arXiv:1602.01412](https://arxiv.org/abs/1602.01412) (2016)
143. Rasley, J., Karanasos, K., Kandula, S., Fonseca, R., Vojnovic, M., Rao, S.: Efficient queue management for cluster scheduling. In: *Proceedings of the Eleventh European Conference on Computer Systems*, p. 36. ACM (2016)
144. Reiss, C., Tumanov, A., Ganger, G.R., Katz, R.H., Kozuch, M.A.: Heterogeneity and dynamics of clouds at scale: Google trace analysis. In: *Proceedings of the Third ACM Symposium on Cloud Computing*, p. 7. ACM (2012)
145. Reiss, C., Wilkes, J., Hellerstein, J.L.: Google cluster-usage traces: format + schema. Technical report, Google Inc., Mountain View, CA, USA, November 2011. <http://code.google.com/p/googleclusterdata/wiki/TraceVersion2>. Accessed 20 Mar 2012
146. Ren, C., Lo, E., Kao, B., Zhu, X., Cheng, R.: On querying historical evolving graph sequences. *Proc. VLDB Endow.* **4**(11), 726–737 (2011)
147. Ren, K., Kwon, Y., Balazinska, M., Howe, B.: Hadoop’s adolescence: an analysis of hadoop usage in scientific workloads. *Proc. VLDB Endow.* **6**(10), 853–864 (2013)
148. Ricciardi, S., Careglio, D., Sole-Pareta, J., Fiore, U., Palmieri, F., et al.: Saving energy in data center infrastructures. In: *2011 First International Conference on Data Compression, Communications and Processing (CCP)*, pp. 265–270. IEEE (2011)
149. Rodriguez, M.G., Leskovec, J., Schölkopf, B.: Structure and dynamics of information pathways in online media. In: *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, Rome, Italy (2013)
150. Rogers, E.M.: *Diffusion of Innovations*. Simon and Schuster, New York (1995)
151. Sastry, N., Yoneki, E., Crowcroft, J.: Buzztraq: predicting geographical access patterns of social cascades using social networks. In: *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems, SNS 2009, Nuremberg, Germany, 31 March 2009*, pp. 39–45 (2009). <http://doi.acm.org/10.1145/1578002.1578009>
152. Satyanarayanan, M.: Pervasive computing: vision and challenges. *IEEE Pers. Commun.* **8**(4), 10–17 (2001)
153. Scellato, S., Mascolo, C., Musolesi, M., Crowcroft, J.: Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In: *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, 28 March–1 April 2011*, pp. 457–466 (2011). <http://doi.acm.org/10.1145/1963405.1963471>
154. Schilit, B.N., et al.: Challenge: ubiquitous location-aware computing and the place lab initiative. In: *Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, pp. 29–35. ACM (2003)



155. Schilit, B.N., Theimer, M.M.: Disseminating active map information to mobile hosts. *IEEE Netw.* **8**(5), 22–32 (1994)
156. Schwarzkopf, M., Konwinski, A., Abd-El-Malek, M., Wilkes, J.: Omega: flexible, scalable schedulers for large compute clusters. In: *Proceedings of the 8th ACM European Conference on Computer Systems*, pp. 351–364. ACM (2013)
157. Sendling, T.C.: *Micromotives and Macrobehavior*. Norton, New York (1978)
158. Shue, D., Freedman, M.J., Shaikh, A.: Performance isolation and fairness for multi-tenant cloud storage. *OSDI* **12**, 349–362 (2012)
159. Sohrabi, S., Tang, A., Moser, I., Aleti, A.: Adaptive virtual machine migration mechanism for energy efficiency. In: *Proceedings of the 5th International Workshop on Green and Sustainable Software*, pp. 8–14. ACM (2016)
160. Song, X., Chi, Y., Hino, K., Tseng, B.L.: Information flow modeling based on diffusion rate for prediction and ranking. In: *Proceedings of the 16th International Conference on World Wide Web*, pp. 191–200. ACM (2007)
161. Sundmaeker, H., Guillemin, P., Friess, P., Woelfflé, S.: Vision and challenges for realising the internet of things. *Clust. Eur. Res. Proj. Internet Things Eur. Commission* **3**(3), 34–36 (2010)
162. Thereska, E., Donnelly, A., Narayanan, D.: Sierra: practical power-proportionality for data center storage. In: *Proceedings of the Sixth Conference on Computer systems*, pp. 169–182. ACM (2011)
163. Torres, R., Finamore, A., Kim, J.R., Mellia, M., Munafo, M.M., Rao, S.: Dissecting video server selection strategies in the YouTube CDN. In: *2011 31st International Conference on Distributed Computing Systems (ICDCS)*, pp. 248–257. IEEE (2011)
164. Tasic, M., et al.: Semantic coordination protocol for LTE and Wi-Fi coexistence. In: *2016 European Conference on Networks and Communications (EuCNC)*, pp. 69–73. IEEE (2016)
165. Traverso, S., Huguenin, K., Trestian, I., Erramilli, V., Laoutaris, N., Papagianaki, K.: TailGate: handling long-tail content with a little help from friends. In: *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, 16–20 April 2012*, pp. 151–160 (2012). <http://doi.acm.org/10.1145/2187836.2187858>
166. Vavilapalli, V.K., et al.: Apache Hadoop YARN: yet another resource negotiator. In: *Proceedings of the 4th Annual Symposium on Cloud Computing*, p. 5. ACM (2013)
167. Vergara, E.J., Nadjm-Tehrani, S.: EnergyBox: a trace-driven tool for data transmission energy consumption studies. In: *Energy Efficiency in Large Scale Distributed Systems - COST IC0804 European Conference, EE-LSDS 2013, Vienna, Austria, 22–24 April 2013, Revised Selected Papers*, pp. 19–34 (2013). [http://dx.doi.org/10.1007/978-3-642-40517-4\\_2](http://dx.doi.org/10.1007/978-3-642-40517-4_2)
168. Verma, A., Pedrosa, L., Korupolu, M., Oppenheimer, D., Tune, E., Wilkes, J.: Large-scale cluster management at Google with Borg. In: *Proceedings of the Tenth European Conference on Computer Systems*, p. 18. ACM (2015)
169. Wang, Z., Zhou, X., Yu, Z., Wang, H., Ni, H.: Quantitative evaluation of group user experience in smart spaces. *Cybern. Syst.: Int. J.* **41**(2), 105–122 (2010)
170. Want, R., Pering, T.: System challenges for ubiquitous & pervasive computing. In: *Proceedings of the 27th International Conference on Software Engineering*, pp. 9–14. ACM (2005)
171. Watts, D.J.: A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci.* **99**(9), 5766–5771 (2002)

172. Yan, Y., Qian, Y., Sharif, H., Tipper, D.: A survey on smart grid communication infrastructures: motivations, requirements and challenges. *IEEE Commun. Surv. Tutor.* **15**(1), 5–20 (2013)
173. Yang, H., Breslow, A., Mars, J., Tang, L.: Bubble-flux: precise online QoS management for increased utilization in warehouse scale computers. In: *ACM SIGARCH Computer Architecture News*, vol. 41, pp. 607–618. ACM (2013)
174. Yang, J., Leskovec, J.: Modeling information diffusion in implicit networks. In: *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM)*, pp. 599–608. IEEE (2010)
175. Zaharia, M., Borthakur, D., Sen Sarma, J., Elmeleegy, K., Shenker, S., Stoica, I.: Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling. In: *Proceedings of the 5th European Conference on Computer Systems*, pp. 265–278. ACM (2010)
176. Zhang, D., Yu, Z., Guo, B., Wang, Z.: Exploiting personal and community context in mobile social networks. In: Chin, A., Zhang, D. (eds.) *Mobile Social Networking. CSS*, pp. 109–138. Springer, New York (2014). [https://doi.org/10.1007/978-1-4614-8579-7\\_6](https://doi.org/10.1007/978-1-4614-8579-7_6)
177. Zhang, X., Tune, E., Hagmann, R., Jnagal, R., Gokhale, V., Wilkes, J.: CPI 2: CPU performance isolation for shared compute clusters. In: *Proceedings of the 8th ACM European Conference on Computer Systems*, pp. 379–391. ACM (2013)
178. Zhou, F., Zhang, L., Franco, E., Mislove, A., Revis, R., Sundaram, R.: WebCloud: recruiting social network users to assist in content distribution. In: *Proceedings of the 11th IEEE International Symposium on Network Computing and Applications*, Cambridge, MA, USA (2012)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

