REVIEW

CrossMark

Feature-rich networks: going beyond complex network topologies

Roberto Interdonato[1]* iD, Martin Atzmueller[2], Sabrina Gaito[3], Rushed Kanawati[4], Christine Largeron[5] and Alessandra Sala[6]

*Correspondence:
roberto.interdonato@cirad.fr
[1]CIRAD - UMR TETIS, Montpellier, France
Full list of author information is available at the end of the article

Abstract

The growing availability of multirelational data gives rise to an opportunity for novel characterization of complex real-world relations, supporting the proliferation of diverse network models such as Attributed Graphs, Heterogeneous Networks, Multilayer Networks, Temporal Networks, Location-aware Networks, Knowledge Networks, Probabilistic Networks, and many other task-driven and data-driven models. In this paper, we propose an overview of these models and their main applications, described under the common denomination of *Feature-rich Networks*, i. e. models where the expressive power of the network topology is enhanced by exposing one or more peculiar *features*. The aim is also to sketch a scenario that can inspire the design of novel feature-rich network models, which in turn can support innovative methods able to exploit the full potential of mining complex network structures in domain-specific applications.

Introduction

Structures built upon great quantities of networked entities, such as computer networks and social networks, have an undeniable central role in our everyday life. The need to study these complex real-world topologies, together with the growing ability to carry out these studies thanks to technological advances, recently made the use of complex network models pervasive in many disciplines such as computer science, physics, social science, as well as in interdisciplinary research environments.

Nowadays, it is straightforward to experience the use of complex networked data, thanks to the fact that collecting multirelational data from the Web is generally a simple and inexpensive task. Just think about the quantity of online social media platforms, crowdsourced data, online knowledge bases, and so on, that can be collected and studied with relatively low effort.

Nevertheless, besides relational data that can be modeled in a network topology, it is easy to recognize a quantity of "extra" features which serve as an inestimable source of information, that can be conveniently embedded in a network, thus enhancing the expressive power of the topology itself. Examples are given by temporal aspects of the data, quantitative and/or qualitative properties of the nodes, different relations between a common set of entities and different existence probabilities.

In this paper, we refer with the term *Feature Rich-Networks* to all the complex network models that expose one or more features in addition to the network topology. Some examples of feature-rich networks, which will be described in the paper, are:

- Attributed graphs, e. g. networks enclosing (vectors of) generic attributes on nodes and edges ("Attributed graphs" section);
- Heterogeneous information networks, e. g. networks modeling heterogeneous node and edge types ("Heterogeneous information networks" section);
- Multilayer networks, e. g. representing different online/offline relations between the same set of users ("Multilayer networks" section);
- Temporal networks, e. g. modeling discrete/continuous time aspects in networked data ("Temporal networks" section);
- Location-aware Networks, e. g. useful for the definition of recommender system (RecSys) applications like itinerary routing and points of interest (PoIs) planning ("Location-aware networks" section);
- Probabilistic networks, e. g. networks modeling uncertain relations, such as sensor networks, or networks inferred from survey data ("Probabilistic networks" section).

Please note that the definition of feature-rich network has been kept intentionally wide and flexible, with the aim to gather under a common denomination a series of network models exhibiting different structures and that were introduced for different needs, but that at the same time show some common characteristics and can lead to similar problems. For the same reason, the overview is not meant to be exhaustive, and other network models may exist which can be referred to as feature-rich ones.

In this paper, we will provide an insight in the current status of research in feature-rich network analysis and mining, describing the main types of feature-rich networks and related applications. The aim is to show how embedding features in complex network models can make it possible to improve solutions to classic tasks (e. g. centrality, community detection, link prediction, information diffusion, and so on) and to focus on domains and research questions that have not been deeply investigated so far.

Attributed graphs

Together with the relational information (i.e., the graph), many data sources may also provide attributes describing the relationships or the entities of the network leading to the notions of a node-attributed graph or an edge-attributed graph, respectively. When the attributes are associated with the relationships, the network can be represented by a weighted graph where the weights, usually used to measure the strength of the tie between the corresponding nodes, are replaced by a vector whose components correspond to attributes characterizing their relation. For instance, in a co-authorship network, the link between two coauthors can be described not only by the total number of their co-publications but also by their dates or by the number for each co-publications sub-type (e. g. conference, journal, etc.). So, a vector can be assigned to the edges to take into account these attributes. Note that in specific cases, alternative network models may be used, such as temporal networks (cf. "Temporal networks" section) for modeling interactions over time or multiplex networks (cf. "Multilayer networks" section) for modeling each attribute by a specific relationship. The concept of (node-) attributed networks refers rather to the case where attributes are assigned to the nodes for describing

the corresponding entities. In a friendship network, e. g. the actors can be described by their genre and their age.

In literature, different definitions have been introduced. A first model has been defined by Zhou et al. (2009), an alternative by Yin et al. (2010):

Definition 1 (Attributed Network - Zhou et al. (2009)) *An attributed network is defined as a graph G = (V, E) where V and E denote sets of nodes and edges; each node $v \in V$ is associated with a is associated with a vector of attributes $(v_j, j \in \{1, .. p\})$*

Definition 2 (Attributed Network with bipartite graph - Yin et al. (2010)) *An attributed network is represented by*

- *a graph G = (V, E) describing the relationships between the entities, and*
- *a bipartite graph $G_a = (V \cup V_a, E_a)$ describing the relationships between the entities and the attributes in such a way that each node v from V is connected to attribute-nodes from V_a.*

The choice of one of these models depends on the type and the number of the features retained to describe the entities of the network: The second definition is more appropriate when few categorical attributes are considered.

In different tasks, taking into account the attributes in addition to the relational information allows to improve the performance of the methods. Thus, attributed networks have been used with success for link prediction, inferring attributes or community detection (Zhou et al. 2010; Yang et al. 2013; Gong et al. 2014; Combe et al. 2015; Atzmueller et al. 2016). However, it is necessary to be careful because structure and attributes may disagree (Peel et al. 2017). Nevertheless, due to the homophily effect and to social influence, they are likely to be aligned, e. g. (McPherson et al. 2001; La Fond and Neville 2010; Mitzlaff et al. 2013; Mitzlaff et al. 2014; Atzmueller and Lemmerich 2018). Consequently, one can hope to benefit from the two sources, notably when one is missing or noisy. Finally it should be mentioned that generators have been recently designed to automatically build attributed networks (Akoglu and Faloutsos 2009; Palla et al. 2012; Kim and Leskovec 2012; Largeron et al. 2017). Such benchmarks are particularly useful for evaluating the performance of algorithms able to handle the two kinds of data.

A well known subcategory of attributed graphs includes the models used for direct organization and modeling of knowledge elements, e. g. given by concepts, their properties and (inter-)relations. Rooted in the theory on semantic networks (Sowa 2006), such models are known as knowledge networks or knowledge graphs (Bizer et al. 2009; Hoffart et al. 2013). In such network structures, data is integrated into a comprehensive knowledge model capturing the relations between concepts and their properties in an explicit way, cf. (Bizer et al. 2009; Hoffart et al. 2013; Ristoski and Paulheim 2016). For instance, entities (concepts) are usually represented as nodes, there can be categories (labels) associated to node, and conceptual relations are given by directed edges between the nodes (Pujara et al. 2013). Following Paulheim (2017), from the point of construction, a knowledge network then mainly describes real world entities and their interrelations. The possible classes and relations can then also be potentially interrelated in an arbitrary way. Knowledge networks can be exploited in many ways, for example, in order to facilitate

modeling, mining, inference, and reasoning. Then, tasks that are supported by knowledge networks include, for example, advanced feature engineering, e. g. (Atzmueller and Sternberg 2017; Wilcke et al. 2017). Furthermore, the constructed knowledge graph can serve as a data integration and exploration mechanism, such that the considered relations and additional information about the contained entities can be utilized by advanced graph mining methods, that work on such feature-rich networks, e. g. by mining the respective attributed graph, e. g. (Atzmueller et al. 2016; Atzmueller et al. 2017). Knowledge graphs thus have a broad range of applications, ranging from knowledge modeling and structuring, cf. (Bizer et al. 2009; Hoffart et al. 2013) to advanced graph mining applications in diverse domains (Ristoski and Paulheim 2016; Wilcke et al. 2017; Atzmueller et al. 2016; Atzmueller and Sternberg 2017).

Heterogeneous information networks

The definition of Heterogeneous Information Network (HIN) models rises from the observation that sophisticated real-world networks can hardly be represented with standard network topologies. Most of real-world connections happen between entities that can be considered as different kinds, and describe different types of relations. A practical example is given by a bibliographic information network, containing entities of type paper, venue and author, where different relation types can connect nodes of different entity types (e. g. authorship between author and paper, publication between paper and venue, and so on) or even nodes of the same type (e. g. coauthorship between authors, citation between papers).

While HINs are a powerful tool to model real-world situations, on the other side the modeling process should be carried out by looking for a good trade-off between homogeneous networks (i. e. all nodes of the same type) and complete heterogeneity (i. e. each node establishes a different entity type), since both extremes would result in a loss of information. For this reason, the authors in Sun and Han (2012) propose a typed, semi-structured heterogeneous network model, defined as follows:

Definition 3 (Heterogeneous Information Network) *An information network is defined as a directed graph $G = (\mathcal{V}, \mathcal{E})$ with an object type mapping function $\tau : \mathcal{V} \rightarrow \mathcal{A}$ and a link type mapping function $\phi : \mathcal{E} \rightarrow \mathcal{R}$, where each object $v \in \mathcal{V}$ belongs to one particular object type $\tau(v) \in \mathcal{A}$, each link $e \in \mathcal{E}$ belongs to a particular relation $\phi(e) \in \mathcal{R}$, and if two links belong to the same relation type, the two links share the same starting object type as well as the ending object type. When the types of objects $|\mathcal{A}| > 1$ or the types of relations $|\mathcal{R}| > 1$, the network is called heterogeneous information network; otherwise, it is a homogeneous information network.*

Given a complex heterogeneous information network, it is necessary to provide its meta level (i. e. schema-level) description for better understanding the object types and link types in the network. Therefore, the concept of network schema is proposed, in order to describe the meta structure of a network (Sun and Han 2012):

Definition 4 (Network Schema) *The network schema, denoted as $T_G = (\mathcal{A}, \mathcal{R})$, is a meta template for a heterogeneous network $G = (\mathcal{V}, \mathcal{E})$ with the object type mapping*

$\tau : \mathcal{V} \rightarrow \mathcal{A}$ *and the link mapping* $\phi : \mathcal{E} \rightarrow \mathcal{R}$, *which is a directed graph defined over object types* $\mathcal{A}$, *with edges as relations from* $\mathcal{R}$.

The network schema of a heterogeneous information network has specified type constraints on the sets of objects and relationships between the objects. These constraints make a heterogeneous information network semi-structured, guiding the exploration of the semantics of the network (Sun and Han 2012). This HIN model has been successfully used for several mining tasks, such us ranking-based clustering combinations (Sun et al. 2009; Sun et al. 2009), transductive and ranking-based classification (Ji et al. 2010; Ji et al. 2011), similarity search (Sun et al. 2011) and relationship prediction (Sun et al. 2012; Deng et al. 2014), and, more recently, learning of object-event embeddings (Gui et al. 2017) and named entity linking (Shen et al. 2018). However, the notion of HIN is general enough to include other network models which are inherently heterogeneous in node and relation types, e. g. networks related to the Internet-of-Things (George and Thampi 2018; Misra et al. 2012; Qiu et al. 2016).

Multilayer networks

Multilayer network models provide a powerful and realistic tool for the analysis of complex real-world network systems, enabling an in-depth understanding of the characteristics and dynamics of multiple, interconnected types of node relations and interactions (Dickison et al. 2016). While they can be seen as a form of HIN (cf. "Heterogeneous information networks" section), the main idea here is to model the different relations which may occur between the same set of entities in different *layers*. The layers can be seen as different interaction contexts, while the participation of an entity to different layers can be seen as a set of different *instances* of the same entity. When the only inter-layer edges (i. e. edges linking instances in different layers) are the coupling edges (i. e. edges linking different instances of the same entity), this model is generally referred to as *Multiplex Network*. As a practical example, in social computing, an individual often has multiple accounts across different social networks. Multilayer networks can be easily used to link distributed user profiles belonging to the same user from multiple platforms, thus enabling the definition of advanced mining tasks, e. g. multilayer community detection (Kim and Lee 2015; Loe and Jensen 2015). Similarly, different layers can be used to model online and offline relations of different types happening in a social network (Gaito et al. 2012; Dunbar et al. 2015), such as followship, like/comment interactions, working relationship, lunch relationship, etc. A multilayer network model which has become very popular in literature is that proposed by Kivela et al. (2014):

Definition 5 (Multilayer Network) *Let* $\mathcal{L} = \{L_1, \ldots, L_\ell\}$ *be a set of layers and* $\mathcal{V}$ *be a set of entities. We denote with* $V_{\mathcal{L}} \subseteq \mathcal{V} \times \mathcal{L}$ *the set containing the entity-layer combinations in which an entity is present in the corresponding layer. The set* $E_{\mathcal{L}} \subseteq V_{\mathcal{L}} \times V_{\mathcal{L}}$ *contains the undirected links between such entity-layer pairs. We hence denote with* $G_{\mathcal{L}} = (V_{\mathcal{L}}, E_{\mathcal{L}}, \mathcal{V}, \mathcal{L})$ *the multilayer network graph with set of nodes* $\mathcal{V}$.

Another multilayer network model, specifically conceived to represent multilayer social networks, is proposed by Magnani and Rossi in Dickison et al. (2016):

Definition 6 (Multilayer Social Network) *Given a set of actors $\mathcal{A}$ and a set of layers $\mathcal{L}$, a multilayer network is defined as a quadruple $G = (\mathcal{A}, \mathcal{L}, V, E)$ where (V, E) is a graph, $V \subseteq \mathcal{A} \times \mathcal{L}$ and $E \subseteq V \times V$.*

In this model the concept of an *Actor* is a model upon the physical user, while the *Nodes* can be seen as the "instances" of the actor/user in different contexts/layers (e. g. accounts on different online social networks, or participation in different offline social networks).

Beyond the social networks domain (Dickison et al. 2016; Perna et al. 2018), multilayer networks have been successfully used to model relations and address mining tasks in different domains, such as airline companies (Cardillo et al. 2013), protein-protein interactions (Bonchi et al. 2014), offline – online networks (Scholz et al. 2013), bibliographic networks (Boden et al. 2012), communication networks (Kim and Lee 2015; Bourqui et al 2016), and remote sensing data (Interdonato et al. 2017).

Temporal networks

Real world phenomena are dynamic by nature, i. e. entities participating in a phenomenon and the interactions between them evolve over time, and each interaction typically happens at a specific time and lasts for a certain duration. *Temporal networks* (Li et al. 2017; Zignani et al. 2014) are the model used to represent these dynamic features in network graphs. Temporal networks have been referred to with different other terms, such as evolving graphs, time-varying graphs, timestamped graphs, dynamic networks, and so on.

Holme and Saramaki (2012) identify two main classes of temporal network, namely *contact sequences* and *interval graphs*. A *contact sequence* network is suitable for cases where there's a set of entities V interacting with each other at certain times, and the durations of the interactions are negligible. Typical systems suitable to be represented as a contact sequence include communication data (sets of e-mails, phone calls, text messages, etc.), and physical proximity data where the duration of the contact is less important (e.g. sexual networks) (Holme and Saramäki 2012). A *contact sequence* network can be defined as follows:

Definition 7 (Contact sequence network) *A contact sequence network $G = (V, C)$ is defined by a set of vertices V with an associated set of contacts C, where each contact $c \in C$ is a triple (i, j, t) where $i, j \in V$ and t is a timestamp denoting a time of contact between i and j. A contact sequence network can be equivalently defined as $G = (V, E, T, f)$, where E is a set of edges, T is a set of non-empty timestamp lists, and $f : E \to T$ is a function associating each edge to its timestamp list such that for each $e \in E$ exists $f(e) = T_e = \{t_1, ..., t_n\}$.*

If the duration of the interactions is considered (i. e. each edge is active at certain time intervals), then the *interval graph* model is more suitable:

Definition 8 (Interval graph) *An interval graph $G = (V, E, T, f)$ is defined by a graph $G = (V, E)$, a set of lists of time intervals T, and a function $f : E \to T$ associating a list of time intervals to each edge $e \in E$, such that $T_e = \{(t_1, t'_1), ..., (t_n, t'_n)\}$, with each couple (t_i, t'_i) denoting the beginning and ending time of a time interval.*

Examples of systems that are natural to model as interval graphs include proximity networks (where a contact can represent that two individuals have been close to each other for some extent of time), seasonal food webs where a time interval represents that one species is the main food source of another at some time of the year, and infrastructural systems like the Internet (Holme and Saramäki 2012). In both cases (i. e. starting from a *contact sequence* network or from an *interval graph*), a static *time aggregated graph* can be derived, where an edge between two nodes *i* and *j* exists if and only if there is at least a contact between *i* and *j*. Temporal networks have been used to address problems in different domains, such as community detection in dynamic social networks (Rossetti et al. 2017), activity pattern analysis of editors (Yasseri et al. 2012), temporal aspects of protein interaction (Han et al. 2004) and gene-regulatory networks (Lèbre et al. 2010), analysis of temporal text networks (Vega and Magnani 2018), analysis of epidemic spreading (Moinet et al. 2018; Onaga et al. 2017) and problems related to mobile devices (Tang et al. 2011; Quadri et al. 2014), just to name a few.

Location-aware networks

As discussed for the time dimension (cf. "Temporal networks" section), in several cases modeling networks from real-world phenomena may require taking into account *spatial* features. The use of location-based (e. g. georeferenced) information is commonly related to specific research fields, e. g. the ones connected to geographical issues and analyses. Nevertheless, in recent years the increasing availability of gps-equipped mobile devices gave rise to the development of location-based social networking (LBSN) services, such as Foursquare, Facebook Places, Google Latitude, Tripadvisor and Yelp. Consequently, several research approaches have been proposed which make use of geographical and spatio-temporal features in social network analysis problems.

Based on the analysis in Bao et al. (2015), in typical cases different types of location-aware networks can be defined, depending on which informations are extracted from the LBSN:

Definition 9 (Location-location graph) *A location-location graph $G = (V, E)$ is a graph where nodes in V represent locations and directed edges in $E \subseteq V \times V$ represent the relation between two locations. The semantic of the relation can be defined in different ways, e. g. distance between the location (i. e. expressed as edge weight), similarity or visits by the same users.*

Definition 10 (User-location graph) *A user-location graph $G = (U, V, E)$ is a bipartite graph where nodes in U represent users, nodes in V represent locations and directed edges in $E \subseteq U \times V$ represent relations between users and locations. The semantic of the relation can be flexible, e. g. may indicate that a user visited or rated a certain location.*

Definition 11 (User-user graph) *A user-user graph $G = (V, E)$ is a graph where nodes in V represent users and directed edges in $E \subseteq V \times V$ represent relations between users. Some typical edge semantics here may be physical distances, friendship on a LBSN, or features derived from users' location histories (e. g. edges may connect users having visited a common location).*

Location-aware networks built upon LSBN data are generally used for Point-of-Interest (POI) recommendation tasks (Bao et al. 2015; Zhang and Chow 2015; Liu 2018), with the aim to combine geographical and social influence in the recommendation process. A location-based Influence Maximization problem is addressed in Zhou et al. (2015), exploiting LSBN to carry out product promotion in a Online to Offline (O2O) business model. A location-aware multilayer network is proposed in Interdonato and Tagarelli (2017), for a POI recommendation task, which integrates location-aware features from a LSBN (Foursquare), geographical features from Google Maps and conceptual features from Wikipedia on different layers.

Networks based on geographical features can also be extracted from remote sensing data, i. e. satellite images. An approached based on evolution graphs is proposed in Guttler et al. (2017), in order to detect spatio-temporal dynamics satellite image time series. Different evolution graphs are produced for particular areas within the study site, which store information about the temporal evolution of a specific geographical area. Then the graph are both studied separately and compared to each other in order to provide a global analysis on the dynamical evolution of the site.

Probabilistic networks

When using networks to model real-world complex phenomena, it is easy to incur in situations where the existence of the relationship between two entities is *uncertain*. The sources of this uncertainty can be manifold, e. g. links may be derived from erroneous or noisy measurements, inferred from probabilistic models (Monti and Boldi 2017), or even intentionally obfuscated for various reasons. A practical example is offered by biological networks representing protein and gene interactions. Since the interactions are observed through noisy and error-prone experiments, link existence is uncertain, and a major part of uncertainty may arise in social networks for reasons related to data collection (e. g. data collected through automated sensors, inferred from anonymized communication data or from self-reporting/logging data (Adar and Ré 2007)), or because the network structure is based on prediction algorithms (e. g. approaches based on link prediction (Liben-Nowell and Kleinberg 2007)), or simply because actual interactions in online and offline social networks are difficult to measure. Similar issues may happen when coping with Temporal (cf. "Temporal networks" section) and Location-aware (cf. "Location-aware networks" section) networks, always due to data collection (von Landesberger et al. 2017; Wunderlich et al. 2017). In specific cases, uncertainty in the link structure may also be intentionally injected in a network for privacy reasons (Boldi et al. 2012).

All these situations can be handled by using *probabilistic network* models, often referred to as *uncertain graphs*, whose edges are labeled with a probability of existence. This probability represents the confidence with which one believes that the relation corresponding to the edge holds in reality (Parchas et al. 2015). A typical probabilistic network, referred to as *Uncertain Graph*, is defined in Parchas et al. (2015):

Definition 12 (Uncertain Graph) *An uncertain graph is defined as a triple* $\mathcal{G} = (V, E, p)$, *where function* $p : E \rightarrow (0, 1]$ *assigns a probability of existence to each edge.*

Following the literature, the authors consider the edge probabilities independent (Potamias et al. 2010; Jin et al. 2011), and assume *possible-world* semantics

(Abiteboul et al. 1987; Dalvi and Suciu 2004). Specifically, the possible-world semantics interprets $\mathcal{G}$ as a set $\{G = (V, E_G)\}_{E_G \subseteq E}$ of $2^{|E|}$ possible deterministic graphs (worlds), each defined by a subset of E. The probability of observing any possible world $G = (V, E_G) \sqsubseteq \mathcal{G}$ is:

$$Pr(G) = \prod_{e \in E_G} p(e) \prod_{e \in E \setminus E_G} (1 - p(e)) \tag{1}$$

Nevertheless, the expressive power enabled by a probabilistic network schema naturally carries with it an explosion in complexity, e. g. the exponential number of possible worlds may even prevent exact query evaluation on the graph. More specifically, even simple queries on deterministic graphs become #P-complete problems on uncertain graphs, and also approximated approaches based on sampling may be too expensive in most cases. To overcome these issues, Parchas et al. propose to create deterministic representative instances of uncertain graphs that maintain the underlying graph properties (Parchas et al. 2015).

Conclusions and future challenges

In this paper, we discussed the main *feature-rich* network models, namely Attributed Graphs, Heterogeneous Information Networks, Multilayer Networks, Temporal Networks, Location-aware Networks and Probabilistic Networks. Table 1 summarizes the main features exposed for nodes and edges for each discussed model. We introduced the term *Feature-rich Network* in order to refer to all the complex network models that expose one or more features in addition to the network topology. We kept the definition intentionally wide, with the aim to gather under a common denomination a series of network models which were introduced for different needs, but that at the same time show some common characteristics and can lead to similar problems. Given the flexibility of the definition, this overview is not meant to be exhaustive, and many other feature-rich network models (e. g. data-driven ones) may exist or may be defined in different domains. The use of Feature-rich Networks can intuitively be perceived as beneficial for most research tasks based on graph data, given the greater quantity of information carried by the network object with respect to classic ones. Nevertheless, their expressive power has not been yet fully valued, therefore there is an emergence for providing insights into how the study of feature-rich network models can pave the way for the definition of domain-specific problems that might not be adequately addressed by classic ones. Moreover, the research community also needs an insight in how correctly handling a richer feature set can lead to the definition of network analysis and mining methods that are able to address classic tasks (e. g. community detection, link prediction, information propagation, and so on), improving upon classic models in terms of results quality, while limiting the impact on their efficiency and scalability. Moreover, the use of feature-rich network models may be beneficial for problems in interdisciplinary research fields. In fact, the interplay among researchers from different fields can help modeling most interesting features, and finding new semantics for well-known network analysis tasks. A (non exhaustive) list of domains which usually cope with interdisciplinary research environments and that would benefit from the use of these models include social sciences, physics, remote sensing, health care support, crime and crisis management.

Table 1 Table summarizing the main features exposed for nodes and edges for the discussed feature-rich network models

Network model	Node features	Edge features
Attributed graph	Attributes vector	Attributes vector
Heterogeneous information network	Object type	Relation type
Multilayer network	Layer	Layer
Temporal network	-	Timestamp/Time interval
Location-aware network	Geolocation	Distance/Visiting/Rating
Probabilistic network	-	Probability of existence

Authors' contributions
RI contributed to "Introduction", "Heterogeneous information networks", "Multilayer networks", "Temporal networks", "Location-aware networks", "Conclusions and future challenges", sections and supervised the writing of the article. MA contributed to sections "Introduction", "Attributed graphs" and "Multilayer networks". CL and RK contributed to "Attributed graphs" and "Multilayer networks" sections. SG and AS contributed to "Temporal networks", "Location-aware networks" and "Probabilistic networks" sections. All authors read and approved the final manuscript.

Authors' information
Roberto Interdonato is a Research Scientist at Cirad, UMR TETIS, Montpellier, France. He was previously a post-doc researcher at University of La Rochelle (France), Uppsala University (Sweden) and at University of Calabria (Italy), where he received his Ph.D. in computer engineering in 2015. His Ph.D. work focused on novel ranking problems in information networks. His research interests include topics in data mining and machine learning applied to complex network analysis (e.g., social media networks, trust networks, semantic networks, bibliographic networks) and to remote sensing. On these topics he has coauthored journal articles and conference papers, organized workshops, presented tutorials at international conferences and developed practical software tools.
Martin Atzmueller is Associate Professor at the Department of Cognitive Science and Artificial Intelligence at Tilburg University as well as Visiting Professor at the Université Sorbonne Paris Cité. He earned his habilitation (Dr. habil.) in 2013 at the University of Kassel, where he also was appointed as adjunct professor (Privatdozent). Further, he received his Ph.D. (Dr. rer. nat.) in Computer Science from the University of Würzburg in 2006. He studied Computer Science at the University of Texas at Austin (USA) and at the University of Würzburg where he completed his MSc in Computer Science. His research areas include data science, data mining network analysis, wearable sensors and big data. He has published more than 200 scientific articles in top venues, e.g., the International Joint Conference on Artificial Intelligence (IJCAI), the European Conference on Machine Learning and Principles and Practice on Knowledge Discovery in Databases (ECML PKDD), the IEEE Conference on Social Computing (SocialCom), the ACM/IEEE International Conference on Advances in Social Networks Analysis and Mining (ASONAM), the ACM International Conference on Information and Knowledge Management (CIKM) and the ACM Conference on Hypertext and Social Media (HT). He is the winner of several Best Paper and Innovation Awards. He regularly acts as PC member of several top-tier conferences and as co-organizer on a number of international workshops, conferences, and tutorials on the topics of data science and network science, in particular on community detection and mining attributed networks. He can be contacted at m.atzmuller@uvt.nl, and his web site is at https://martin.atzmueller.net. Contact info: Tilburg University, Department of Cognitive Science and Artificial Intelligence, Warandelaan 2, 5037 AB Tilburg, Netherlands, Tel: +31-(0)13 466 4736, m.atzmuller@uvt.nl
Sabrina Gaito received a degree in Physics in 1996, a Master Degree in Material Science in 1998 and a Ph.D. in Applied Mathematics in 2002 from the University of Milano, Italy. She is currently Assistant Professor at the same University, where she teaches Social Media Mining and Computer Networks. Her research activity takes place within both the network science, with a focus on complex network theory applied to social networking, human mobility and behaving, and the network technology, with a focus on ad hoc networks and mobile applications.
Rushed Kanawati is an associate professor in computer science at University Paris 13, France, researcher at LIPN CNRS UMR 7030. He is also head of computer networks department at the technological institute, Villetaneuse. He has received a PhD degree in computer science from INPG France in 1998. He joined the INRIA as an expert engineer where he worked on designing and implementing recommender systems. His recent research work covers various topics such as link prediction and community detection in complex networks as well as multiplex and attributed network analysis. He is author of more than 150 papers in national and international venues. He is regularly involved in organizing conferences, workshops and tutorials mainly in the area of complex network analysis. More information can be found at his web site : http://lipn.fr/~kanawati Contact info: University Paris 13 - LIPN CNRS UMR 7030. 99 Av. J-B Clément 93430 Villetaneuse, Tel:+33-(0)1 49404077, rushed.kanawati@lipn.univ-paris13.fr
Christine Largeron is a full professor in computer science. She received a Ph.D in Computer Science from Claude Bernard University (Lyon – France) in 1991. She is Professor at Jean Monnet University (France) since 2006 and, she is the head of the Data Mining and Information Retrieval group of the Hubert Curien Laboratory. Her research interests focus

on machine learning, data mining, information retrieval, text mining, social mining, network analysis. She regularly acts as PC member of several top-tier conferences and as co-organizer on a number of international workshops and conferences. More information can be found on her web site : https://perso.univ-st-etienne.fr/largeron/ Contact info: Laboratoire Hubert Curien - UMR CNRS 5516 (LHC), Jean Monnet University, Saint-Etienne 18, rue Benoit Lauras 42000 Saint-Etienne, Tel: +33-(0)4 77 91 57 56 (57 80) Christine.Largeron@univ-st-etienne.fr

Alessandra Sala, head of Analytics Research Group in Bell-Labs, is the global lead of the analytics research program in Bell-Labs. She is responsible to deliver breakthrough research assets to create new marker opportunities and technology that has the potential to change our human lives. In her prior appointment, she was the technical manager of the Data Analytics and Operations Research group in Bell Labs Ireland. Before that, she held a research associate position in the Department of Computer Science at University of California Santa Barbara. During this appointment, she was a key contributor of several funded proposals from National Science Foundation in USA and her research was awarded with the Cisco Research Award in 2011. She focused her research on modeling massive graphs with an emphasis on mitigating privacy threats for Online Social Network users. Before that, she worked for two years as post-doctoral fellow with the CurrentLab research group led by Prof. Ben Y.Zhao. Before UCSB, she completed her Ph.D in Computer Science at University of Salerno, Italy. Her research focus lies on distributed algorithms, data analytics and complexity analysis with an emphasis on graph algorithms and recently AI, machine learning and deep learning. In her previous research she has developed efficient distributed routing algorithms that support robust and flexible application level services such as scalable search, flexible data dissemination, and reliable anonymous communication. She was a Track Chair of WWW 2016, the general chair of ACM COSN 2014 and has served on the TPC of several networking and data mining conferences as IEEE INFOCOM, WWW, SEA, ICWSM, P2P etc. In 2015 she was awarded Distinguished Member of the IEEE INFOCOM Technical Program Committee.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]CIRAD - UMR TETIS, Montpellier, France. [2]Tilburg University, Tilburg, The Netherlands. [3]Università degli Studi di Milano, Milan, Italy. [4]Université Sorbonne Paris Cité, Paris, France. [5]Laboratoire Hubert Curien, Université de Lyon, Université Jean Monnet, Saint-Etienne, France. [6]Nokia Bell Labs, Dublin, Ireland.

References

Abiteboul S, Kanellakis PC, Grahne G (1987) On the representation and querying of sets of possible worlds. In: Proceedings of the Association for Computing Machinery Special Interest Group on Management of Data 1987 Annual Conference, San Francisco, CA, USA, May 27-29, 1987. pp 34–48

Adar E, Ré C (2007) Managing uncertainty in social networks. IEEE Data Eng Bull 30(2):15–22

Akoglu L, Faloutsos C (2009) Rtg: a recursive realistic graph generator using random typing. Data Min Knowl Disc (DMKD) 19(2):194–209

Atzmueller M, Doerfel S, Mitzlaff F (2016) Description-Oriented Community Detection using Exhaustive Subgroup Discovery. Inf Sci 329:965–984. Publisher: Elsevier, United States

Atzmueller M, Kloepper B, Mawla HA, Jäschke B, Hollender M, Graube M, Arnu D, Schmidt A, Heinze S, Schorer L, Kroll A, Stumme G, Urbas L (2016) Big Data Analytics for Proactive Industrial Decision Support: Approaches & First Experiences in the Context of the FEE Project. atp edition 58(9)

Atzmueller M, Lemmerich F (2018) Homophily at Academic Conferences. In: Proc. WWW 2018 (Companion). ACM Press, New York

Atzmueller M, Schmidt A, Kloepper B, Arnu D (2017) HypGraphs: An Approach for Analysis and Assessment of Graph-Based and Sequential Hypotheses. In: New Frontiers in Mining Complex Patterns. Postproceedings NFMCP 2016, volume 10312 of LNAI. Springer, Berlin/Heidelberg

Atzmueller M, Sternberg E (2017) Mixed-Initiative Feature Engineering Using Knowledge Graphs. In: Proc. 9th International Conference on Knowledge Capture (K-CAP). ACM Press, New York

Bao J, Zheng Y, Wilkie D, Mokbel MF (2015) Recommendations in location-based social networks: a survey. GeoInformatica 19(3):525–565

Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S (2009) Dbpedia-a crystallization point for the web of data. Web Semant Sci Serv Agents World Wide Web 7(3):154–165

Boden B, Günnemann S, Hoffmann H, Seidl T (2012) Mining coherent subgraphs in multi-layer graphs with edge labels. In: Proc. ACM KDD. ACM Press, New York. pp 1258–1266

Boldi P, Bonchi F, Gionis A, Tassa T (2012) Injecting uncertainty in graphs for identity obfuscation. PVLDB 5(11):1376–1387

Bonchi F, Gionis A, Gullo F, Ukkonen A (2014) Distance oracles in edge-labeled graphs. In: Proc. EDBT. pp 547–558

Bourqui R, Ienco D, Sallaberry A, Poncelet P (2016) Multilayer graph edge bundling. In: Proc. PacificVis. IEEE Computer Society, Washington, D.C. pp 184–188

Cardillo A, Gomez-Gardenes J, Zanin M, Romance M, Papo D, del Pozo F, Boccaletti S (2013) Emergence of network features from multiplexity. Sci Rep 3:1344

Combe D, Largeron C, Géry M, Egyed-Zsigmond E (2015) I-louvain: An attributed graph clustering method. In: Advances in Intelligent Data Analysis XIV - 14th International Symposium, IDA 2015, Saint Etienne, France, October 22-24, 2015, Proceedings. Springer, Berlin/Heidelberg. pp 181–192

Dalvi NN, Suciu D (2004) Efficient query evaluation on probabilistic databases. In: (e)Proceedings of the Thirtieth International Conference on Very Large Data Bases, Toronto, Canada, August 31 - September 3 2004. Morgan Kaufmann, Burlington. pp 864–875

Deng H, Han J, Li H, Ji H, Wang H, Lu Y (2014) Exploring and inferring user-user pseudo-friendship for sentiment analysis with heterogeneous networks. Stat Anal Data Min 7(4):308–321

Dickison ME, Magnani M, Rossi L (2016) Multilayer social networks. Cambridge University Press, Cambridge

Dunbar RIM, Arnaboldi V, Conti M, Passarella A (2015) The structure of online social networks mirrors those in the offline world. Soc Networks 43:39–47

Gaito S, Rossi GP, Zignani M (2012) Facencounter: Bridging the gap between offline and online social networks. In: Eighth International Conference on Signal Image Technology and Internet Based Systems, SITIS 2012, Sorrento, Naples, Italy, November 25-29, 2012. IEEE Computer Society, Washington, D.C. pp 768–775

George G, Thampi SM (2018) A graph-based security framework for securing industrial iot networks from vulnerability exploitations. IEEE Access 6:43586–43601

Gong NZ, Talwalkar A, Mackey L, Huang L, Shin ECR, Stefanov E, (Runting) Shi E, Song D (2014) Joint link prediction and attribute inference using a social-attribute network. ACM Trans Intell Syst Technol 5(2):27:1–27:20

Gui H, Liu J, Tao F, Jiang M, Norick B, Kaplan LM, Han J (2017) Embedding learning with events in heterogeneous information networks. IEEE Trans Knowl Data Eng 29(11):2428–2441

Guttler F, Ienco D, Nin J, Teisseire M, Poncelet P (2017) A graph-based approach to detect spatiotemporal dynamics in satellite image time series. ISPRS J Photogramm Remote Sens 130:92–107

Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature

Hoffart J, Suchanek FM, Berberich K, Weikum G (2013) Yago2: A spatially and temporally enhanced knowledge base from wikipedia. Artif Intell 194:28–61

Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519(3):97–125

Interdonato R, Tagarelli A (2017) Personalized recommendation of points-of-interest based on multilayer local community detection. In: Proc. Social Informatics - 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15 2017, Proceedings, Part I. Springer, Berlin/Heidelberg. pp 552–571

Interdonato R, Tagarelli A, Ienco D, Sallaberry A, Poncelet P (2017) Local community detection in multilayer networks. Data Min Knowl Discov 31(5):1444–1479

Ji M, Han J, Danilevsky M (2011) Ranking-based classification of heterogeneous information networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011. ACM Press, New York. pp 1298–1306

Ji M, Sun Y, Danilevsky M, Han J, Gao J (2010) Graph regularized transductive classification on heterogeneous information networks. In: Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24 2010, Proceedings, Part I. Springer, Berlin/Heidelberg. pp 570–586

Jin R, Liu L, Aggarwal CC (2011) Discovering highly reliable subgraphs in uncertain graphs. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011. ACM Press, New York. pp 992–1000

Kim J, Lee J-G (2015) Community detection in multi-layer graphs: A survey. SIGMOD Record 44(3):37–48

Kim M, Leskovec J (2012) Multiplicative attribute graph model of real-world networks. Internet Math 8(1-2):113–160

Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Mutilayer networks. J Complex Networks 2(3):203–271

La Fond T, Neville J (2010) Randomization tests for distinguishing social influence and homophily effects. In: Proceedings of the 19th international conference on World wide web. ACM, New York. pp 601–610

Largeron C, Mougel P-N, Benyahia O, Zaïane OR (2017) Dancer: dynamic attributed networks with community structure generation. Knowl Inf Syst 53(1):109–151

Lèbre S, Becq J, Devaux F, Stumpf MPH, Lelandais G (2010) Statistical inference of the time-varying structure of gene-regulation networks. BMC Syst Biol 4(1):130

Li A, Cornelius SP, Liu Y-Y, Wang L, Barabási A-L (2017) The fundamental advantages of temporal networks. Science 358(6366):1042–1046

Liben-Nowell D, Kleinberg JM (2007) The link-prediction problem for social networks. JASIST 58(7):1019–1031

Liu S (2018) User modeling for point-of-interest recommendations in location-based social networks: The state of the art. Mob Inf Syst 2018:7807461:1–7807461:13

Loe CW, Jensen HJ (2015) Comparison of communities detection algorithms for multiplex. Physica A 431:29–45

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. Annu Rev Sociol 27(1):415–444

Misra S, Barthwal R, Obaidat MS (2012) Community detection in an integrated internet of things and social network architecture. In: 2012 IEEE Global Communications Conference (GLOBECOM). IEEE Computer Society, Washington, D.C. pp 1647–1652

Mitzlaff F, Atzmueller M, Hotho A, Stumme, G (2014) The Social Distributional Hypothesis. J Soc Netw Anal Min 4(216):1–14

Mitzlaff F, Atzmueller M, Stumme G, Hotho A (2013) Semantics of User Interaction in Social Media. In: Ghoshal G, Poncela-Casasnovas J, Tolksdorf R (eds). Complex Networks IV, volume 476 of Studies in Computational Intelligence. Springer, Heidelberg

Moinet A, Pastor-Satorras R, Barrat A (2018) Effect of risk perception on epidemic spreading in temporal networks. Phys Rev E 97:012313

Monti C, Boldi P (2017) Estimating latent feature-feature interactions in large feature-rich graphs. Internet Math:2017

Onaga T, Gleeson JP, Masuda N (2017) Concurrency-induced transitions in epidemic dynamics on temporal networks. Phys Rev Lett 119:108301

Palla K, Knowles DA, Ghahramani Z (2012) An infinite latent attribute model for network data. In: Proceedings of the 29th International Conference on Machine Learning (ICML). Omnipress, USA. pp 1607–1614

Parchas P, Gullo F, Papadias D, Bonchi F (2015) Uncertain graph processing through representative instances. ACM Trans Database Syst 40(3):20:1–20:39

Paulheim H (2017) Knowledge graph refinement: A survey of approaches and evaluation methods. Semant web 8(3):489–508

Peel L, Larremore DB, Clauset A (2017) The ground truth about metadata and community detection in networks. Sci Adv 3(5). American Association for the Advancement of Science

Perna D, Interdonato R, Tagarelli A (2018) Identifying users with alternate behaviors of lurking and active participation in multilayer social networks. IEEE Trans Comput Soc Syst 5(1):46–63

Potamias M, Bonchi F, Gionis A, Kollios G (2010) k-nearest neighbors in uncertain graphs. PVLDB 3(1):997–1008

Pujara J, Miao H, Getoor L, Cohen W (2013) Knowledge graph identification. In: International Semantic Web Conference. Springer, Berlin/Heidelberg. pp 542–557

Qiu T, Luo D, Xia F, Deonauth N, Si W, Tolba A (2016) A greedy model with small world for improving the robustness of heterogeneous Internet of Things. Comput Netw 101:127–143

Quadri C, Zignani M, Capra L, Gaito S, Rossi GP (2014) Multidimensional human dynamics in mobile phone communications. PLoS ONE 9(7):1–12

Ristoski P, Paulheim H (2016) Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey. Web Semant 36:1–22

Rossetti G, Pappalardo L, Pedreschi D, Giannotti F (2017) Tiles: an online algorithm for community discovery in dynamic social networks. Mach Learn 106(8):1213–1241

Scholz C, Atzmueller M, Barrat A, Cattuto C, Stumme G (2013) New Insights and Methods For Predicting Face-To-Face Contacts. In: Kiciman E, Ellison NB, Hogan B, Resnick P, Soboroff I (eds). Proc. 7th Intl. AAAI Conference on Weblogs and Social Media. AAAI Press, Palo Alto

Shen W, Han J, Wang J, Yuan X, Yang Z (2018) SHINE+: A general framework for domain-specific entity linking with heterogeneous information networks. IEEE Trans Knowl Data Eng 30(2):353–366

Sowa JF (2006) Semantic networks. Encycl Cogn Sci. https://doi.org/10.1002/0470018860.s00065

Sun Y, Han J (2012) Mining Heterogeneous Information Networks: Principles and Methodologies. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers

Sun Y, Han J, Zhao P, Yin Z, Cheng H, Wu T (2009) RankClus: integrating clustering with ranking for heterogeneous information network analysis. In: Proc. Int. Conf. on Extending Database Technology (EDBT). ACM Press, New York. pp 565–576

Sun Y, Yu Y, Han J (2009) Ranking-based clustering of heterogeneous information networks with star network schema. In: Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD). ACM Press, New York. pp 797–806

Sun Y, Han J, Aggarwal CC, Chawla NV (2012) When will it happen?: relationship prediction in heterogeneous information networks. In: Proceedings of the Fifth International Conference on Web Search and Web Data Mining, WSDM 2012, Seattle, WA, USA, February 8-12, 2012. ACM, New York. pp 663–672

Sun Y, Han J, Yan X, Yu PS, Wu T (2011) Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. PVLDB 4(11):992–1003

Tang JK, Mascolo C, Musolesi M, Latora V (2011) Exploiting temporal complex network metrics in mobile malware containment. In: 12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WOWMOM 2011, Lucca, Italy, 20-24 June, 2011. IEEE Computer Society, Washington, D.C. pp 1–9

Vega D, Magnani M (2018) Foundations of temporal text networks. Appl Netw Sci 3(1):25:1–25:26

von Landesberger T, Bremm S, Wunderlich M (2017) Typology of uncertainty in static geolocated graphs for visualization. IEEE Comput Graph Appl 37(5):18–27

Wilcke X, Bloem P, de Boer V (2017) The Knowledge Graph as the Default Data Model for Learning on Heterogeneous Knowledge. Data Sci 1(1-2):39–57

Wunderlich M, Ballweg K, Fuchs G, von Landesberger T (2017) Visualization of delay uncertainty and its impact on train trip planning: A design study. Comput Graph Forum 36(3):317–328

Yang J, McAuley J, Leskovec J (2013) Community Detection in Networks with Node Attributes. In: 2013 IEEE 13th International Conference on Data Mining. IEEE Computer Society, Washington, D.C. pp 1151–1156

Yasseri T, Sumi R, Kertész J (2012) Circadian patterns of wikipedia editorial activity: A demographic analysis. PLoS ONE 7:1–8

Yin Z, Gupta M, Weninger T, Han J (2010) Linkrec: A unified framework for link recommendation with user attributes and graph structure. In: Proceedings of the 19th International Conference on World Wide Web, *WWW '10*. ACM Press, New York. pp 1211–1212

Zhang J-D, Chow C-Y (2015) Point-of-interest recommendations in location-based social networks. SIGSPATIAL Special 7(3):26–33

Zhou T, Cao J, Liu B, Xu S, Zhu Z, Luo J (2015) Location-based influence maximization in social networks. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015. ACM Press, New York. pp 1211–1220

Zhou Y, Cheng H, Yu JX (2009) Graph clustering based on structural/attribute similarities. Proc VLDB Endow 2(1):718–729

Zhou, Y, Cheng H, Yu JX (2010) Clustering large attributed graphs: An efficient incremental approach. In: 2010 IEEE International Conference on Data Mining. IEEE Computer Society, Washington, D.C. pp 689–698

Zignani M, Gaito S, Rossi GP, Zhao X, Zheng H, Zhao BY (2014) Link and triadic closure delay: Temporal metrics for social network dynamics. In: Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014. The AAAI Press, USA