Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 52 (2015) 226 – 233

Procedia Computer Science

The 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015)

Descriptive Modeling of Social Networks

Erick Stattner[a], Martine Collard[b]

[a]*erick.stattner@univ-ag.fr, LAMIA Laboratory, University of the French West Indies and Guiana, France*
[b]*martine.collard@univ-ag.fr, LAMIA Laboratory, University of the French West Indies and Guiana, France*

Abstract

These last years, many analysis methods have been proposed to extract knowledge from social networks. As for the traditional data mining domain, these network-based approaches can be classified according to two main families. The approaches based on predictive modelling, which encompass the techniques that analyse current and historical facts to make predictive assumptions about future or unknown events. The approaches based on descriptive modelling, which cover the set of techniques that aim to summarize the data by identifying some relevant features in order to describe how things organize and actually work. In this paper, we review the main descriptive modelling methods of social networks and show for each of them the resulting useful knowledge on a running example. We particularly emphasize on the most recent methods that combine information available on both the network structure and the node attributes in order to provide original description models taking into account the context.

Peer-review under responsibility of the Conference Program Chairs
Keywords: complex network, data mining, social network mining, descriptive modeling, clustering, searching for patterns

1. Introduction

The new "*science of networks*"[2,45,4] is an open and emerging discipline that studies phenomena of the world by focusing on relationships maintained between objects and not on objects as independent entities. While pioneering works have essentially exploited the measures coming from the graph theory[9,5,27,21], new approaches, known as "*social network mining*" (or more simply "*link mining*") aim to take advantages both of the traditional data mining area and network structures. As stated by Getoor and Diehl[15], this area refers to "*data mining techniques that explicitly consider links when building predictive or descriptive models of linked data*". Thus, numerous social network mining methods have been proposed for extracting various kinds of knowledge from social networks.

As for the traditional data mining area, the social network mining domain addresses a large variety of tasks such as classification[23], clustering[11], search for frequent patterns[6] or the link prediction[25]. The inherently descriptive or

* Corresponding author. Tel.: +590-590-483-431
E-mail address: erick.stattner@univ-ag.fr

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2015.05.505

predictive nature of these approaches naturally allows to classify these approaches according to two families: (i) The approaches based on *predictive modelling*, which encompass the techniques that analyse current and historical facts to make predictive assumptions about future or unknown events. (ii) The approaches based on *descriptive modelling*, which cover the set of techniques that aim to summarize the data by identifying some relevant features in order to describe how things organize and actually work.

In this paper, we review the main network-based descriptive models and we describe, for each of them, the potential useful knowledge that can be obtained through a running example. We particularly emphasize on the most recent methods that combine various sources of information coming from the network (for instance structure and node attributes) in order to provide original descriptive models taking into account the context. This work is motivated by the fact that more and more efforts are currently made to design new and innovative descriptive models suitable to the context, which have to be able to answer to specific questions. For instance, the extraction of communities in social networks, which has traditionally been driven by notion of modularity is now extended to consider more homogeneous groups in terms of social links and attributes.

Thus, we identify five main families of methods focusing on descriptive patterns from social networks. (i) link-based clustering that searches for densely connected groups of nodes, (ii) hybrid clustering that consider simultaneously structure and node attributes to identify clusters, (iii) frequent sub-graphs discovery that focuses on the extraction of sub-structures that frequently occur in the network, (iv) network based conceptual analysis that aim to identify the groups of nodes sharing common attributes and (v) frequent conceptual links extraction that provide synthetic and semantic representations on the groups of nodes the most connected of the network. In this paper, we review each family of methods and for each of them we propose an instance of model built from the reference network depicted on Figure 1.

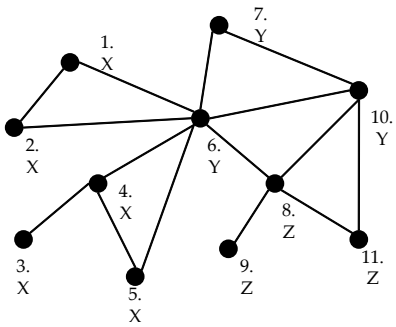

Fig. 1. Reference network

The rest of the paper is organised as follows. Sections 2 to 6 are respectively devoted to the families of method (i) to (v) presented above and Section 7 concludes the paper.

2. Link-based clustering

Link-based clustering (LBC) is one of the problems that has received the most attention from the scientist community. It is also known in the literature as being the problem of *community identification*. More precisely, link-based clustering covers all the methods that aim to group the nodes on a criterion considering links between nodes[43,50].

The objective of LBC techniques is to decompose the network in several connected components, called *community*, so that the nodes in each component have a high density of links while nodes in different component have a lower link density. It is said that the network have a community structure when such groups can be found in the network. It is currently observed that various real world-network have strong community structure[17].
An another definition of a community in the context of social networks, is that two given nodes are more likely to be connected if they belong to the same community. The quality of the network clustering is measured by the so-called modularity measure introduced by M. E. J. Newman[29].

Figure 2 shows an example of the communities that can be obtained from the reference network of Figure 1. Each group of nodes inside dashed circle represents a community. As stated, the network structure is the single information

used to compute the clusters. Consequently, nodes inside a community are strongly connected but can have very different attributes. For example, the community containing nodes 1 and 2 and those containing nodes 3, 4 and 5 share the attribute 'X' but are identified in separate communities.

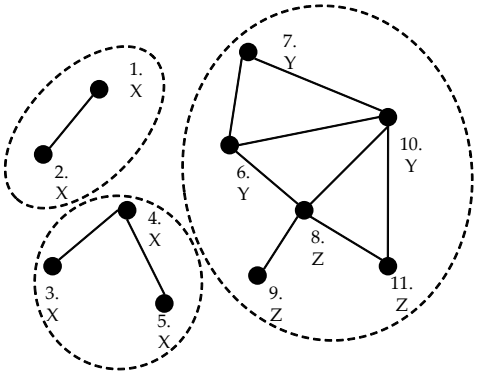

Fig. 2. Communities extracted from the reference network

The principle of the algorithms is to maximize the amount of the intra-community links, while minimize the amount of inter-community ones. Three families of algorithms are commonly distinguished:

- (i) *Agglomerative algorithms*[33,36] refer to bottom-up approaches that iteratively merge nodes or groups of nodes according to a similarity score.
- (ii) *Divisive algorithms*[16,30,34] cover top-down approaches in which the initial network is iteratively fragmented into more and more smaller components by removing the links between groups of nodes having low similarity scores.
- (iii) *Optimisation-based algorithms*[47,28,3], include all techniques that rely on the maximisation of an objective function.

As clustering in networks is an area that has been the subject of a large amount of contributions, other classifications can also be found in the literature. For instance, we can also distinguish methods based on statistical inference or methods based on the concept of communities overlapping, which assume that a node can belong to several communities. A more complete review of the link-based clustering methods can be found in[11].

3. Hybrid clustering

Many scientific and commercial applications need clusters that are more complicated than groups of nodes densely connected. Indeed, in several applications, networks are modeled by links and nodes that may have various kinds of associated attributes. Such networks are called "*information networks*" or "*network with content*"[8]. For example, in a telecommunication network, consumers (nodes) may be identified by attributes such as the *age*, the *type of package*, the *job status*, etc.
Thus in[49] Yang et al. state that by considering only the network structure "an algorithm may fail to account for important structure in the data". That is the reason why numerous works have attempted to take into account this contextual information to better describe the entities and their relations.

Hybrid clustering is one of these new approaches, that consider node attributes during the network clustering process[44,50]. For include the attributes of the nodes in the network clustering task, the definition of a "*cluster*" has been adapted as follows: "*a densely connected group of nodes with homogeneous attributes values*"[51]. In concrete terms, the objective of this new kind of approaches is to partition the network by seeking a balance between the similarity of the structure and the similarity of the attributes so that the nodes with common attributes are grouped in the same partition and nodes into a same partition are densely connected. This kind of approach provides a more semantic partitioning of the network which is necessarily suited to the context.

Figure 3 shows example of hybrid clusters obtained from the reference network. As you can see, hybrid clusters highlight densely connected groups nodes in which nodes have homogeneous attributes. For example, the community formed by the nodes 6, 7 and 10 is both densely connected and share the common attribute 'Y'.

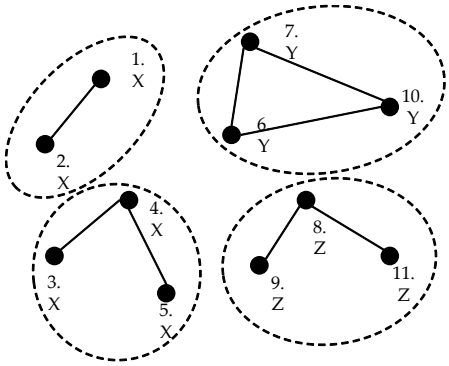

Fig. 3. Hybrid clustering from reference network

Although this approach is relatively recent, many algorithms can already be found in the literature. For example, we can cite the *SA-Cluster* algorithm proposed by Zhou et al.[51], one of the first algorithms based on both structural and attribute similarities through a unified distance measure. We can also cite the ToTem algorithm[7] that extend the classical network clustering method proposed by Blondel et al.[3]. More recently, Yang et al. have proposed the CESNA algorithm[49], for detecting overlapping communities in networks with node attributes. A classification of these approaches, according to two dimensions, can be found in[49].

4. Frequent sub-graphs discovery

Always with the aim to take into account contextual information, a research axis has been focused on the frequency of patterns involving node information and network structure. Indeed, as for the traditional data mining area, some works have been carried out on the extraction of frequent patterns in social networks. The first challenge for these approaches was to define "*a pattern*" in the context of the networks. Thus, the most widely used definition of a pattern is that of a *connected subgraph*[15].

Thus the techniques that focus on the search for frequent patterns in social networks aim to identify the sub-graphs that occur frequently in a single very large network or in a database of networks, according to a minimum support threshold. These approaches assume that a label is associated to each node of the network. In this context, frequent patterns are all the sub-graphs linking a subset of labels that are found frequently enough in the whole network.
A classical example of the use of techniques for the discovery of frequent sub-graphs discovery concerns the basket of items. Indeed, let us consider the network obtained from a basket of items in which nodes correspond to the items and all items are connected to each other when they belong to the same basket, namely when they are purchased together. Once such networks are created for all baskets of consumers, sub-graphs occurring frequently will form frequent patterns in the traditional sense.

Figure 4 shows some frequent sub-graphs extracted from the reference network of Figure 1. For example, the frequent sub-graph composed of labels 'X' and 'Y' is extracted from the initial network. This means that there many links between the nodes that have label 'X' and those that have label 'Y'.

The techniques for searching for frequent sub-graphs in a social network traditionally involve two key steps: (i) *a candidate generation step*, in which candidate sub-graphs potentially frequent in the whole network are generated, and (ii) *an evaluation step* that checks if two networks are isomorphic in order to measure how much frequent the candidates are, according to a frequency threshold. Obviously, these steps are not carried out separately.

Two families of algorithms are commonly distinguished:

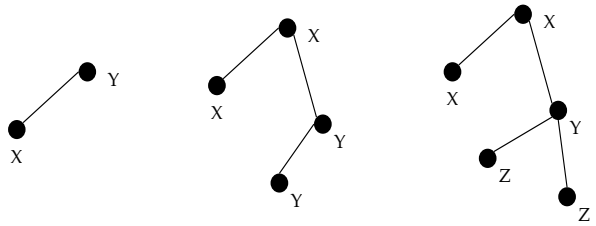

Fig. 4. Frequent sub-graphs from reference network

- (i) *Apriori-based algorithms*[22], that cover all the techniques exploiting the principles of the well-known Apriori algorithm[1] for generating candidates sub-graphs of size t from frequent sub-graphs of size $t - 1$. This kind of approaches starts by searching all frequent sub-graphs among all subgraphs of size 1.
- (ii) *Pattern growth algorithms*[31] refer to approaches that increase a frequent sub-graph of size t to obtain candidate sub-graphs of $t + 1$ by adding a new edge in every possible directions. This kind of approaches must treat the problem of the similar sub-graphs that can be generated at several iterations.

Other classifications suggest to distinguish the sub-graphs discovery methods according to they are interested in the search for frequent subgraphs in a single network or in databases of networks[6].

In the family of Apriori-based approaches, we can cite the *AGM* algorithm proposed by Inokuchi et al.[20] for minimizing both storage and computation. We can also cite the *FSG* algorithm proposed by Kuramochi and Karypis[22] that performs the search for sub-graphs in very big network databases. Regarding the pattern-growth approaches, we can cite the *gSpan* algorithm proposed by Yan and Han[48] that aims to improve the cost of the candidate generation phase by optimizing the discovery process of duplicate structures or the *SPIN* algorithm proposed by Huan et al.[19] that focus on maximal patterns, namely patterns that are not included in any other frequent patterns. More recently some approaches devoted to specific structures, like trees, have been proposed[32]. Good reviews of the frequent patterns mining in networks can be found in[6] and in[18] (see section devoted to structural patterns).

5. Network-based conceptual analysis

Once again, in the context of the simultaneous consideration of node attributes and network structure, some very recent work have attempted to apply the principles of formal concept analysis (FCA)[13] to social networks. These works aim to address the world in terms of objects and attributes by highlighting "*concepts*" composed of sets of nodes (the objects) with their common attributes.

Formal concept analysis is a mathematical approach for data analysis introduced in 1984 by Rudolf Wille[46] that aims to extract knowledge from structured data. The founding principle is to group the objects into classes according to the properties they share. Thus, a couple *(Objets, sharing propoperties)* is called *a concept* and the set of concepts forms a *Galois lattice*, also called *concept lattice*[26,23]. Typical applications of this kind of approach can be found in the field of conceptual clustering[10].

In the context of social networks, recent works conducted on the conceptual analysis applied to networks have considered the network nodes as objects to group and nodes with which they are connected (their neighbors) as the attributes[37,35]. By this way, social networks-based conceptual analysis consists in extracting from the network the groups of node having common neighbors. It is important to note that the nodes of the network can be identified in several concepts since groups may overlap. When personal information are available on the network nodes, the concept lattice can also be built with the individual attributes of nodes.

Figure 5 depicts the concept lattice obtained from the reference network of Figure 1. On such a structure, it is easy to observe the groups nodes sharing common attributes and the inclusive relationships between these groups. Let us specify that we are here in a very simple case, since nodes have only a single attribute, which explain that no concept is identified in the second level of the lattice.

Thus, network-based conceptual analysis has been used to address clustering tasks in social networks[26]. Other works have focused on various problems related to the visualisation[37], the classification[23] or the identification of

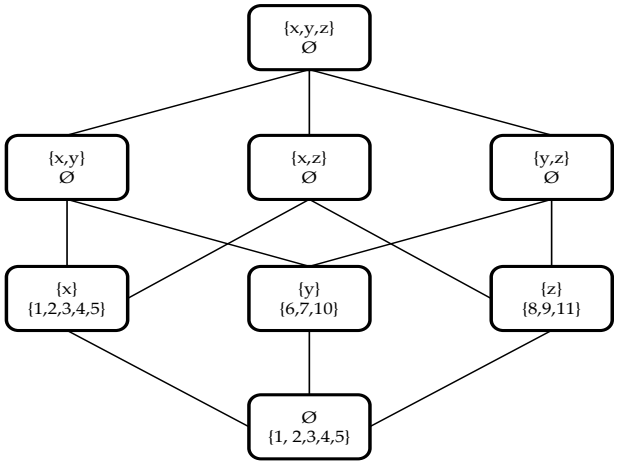

Fig. 5. Concept lattice obtained from the reference network

relevant patterns[35].
For instance, Snasel et al.[38] focused on the problem of visualization and propose a solution that reduces the links between objects in order to optimize the generation of concept lattice. In[14] Gaume et al. propose an approach using FCA for extracting communities in bipartite networks.

A challenging problem concerns to the size of the concept lattice, which is known to grow exponentially with the size of the network, which make it difficult to obtain and may quickly make it uninterpretable. Thus, to avoid the generation and the manipulation of the whole lattice, Le Grand et al.[24] propose several measures to characterize a node. The proposed approach can be extended to the whole dataset in order to characterize the network in terms of homogeneity/heterogeneity and identifying the most significant elements.

6. Frequent conceptual links

In previous sections, we have observed how various kinds of methods have been proposed to extract new and original patterns from social networks. However, the descriptive models extracted do not allow to answer to all the questions that may emerge of the study of social networks. For instance: (i) What are the most connected network node groups? or (ii) What are the attributes the most frequently found in connection into the network? (iii) Are there a groups of individuals most connected than the other ones?

Thus frequent conceptual links (FCL) refers to a new approach[40] that combines information on both network structure and node properties to provide knowledge on the groups of nodes the most connected in the network, in which each group is defined as sets of nodes sharing common attributes. "*Conceptual*" here means that such a link is not a real social link, but represents a set of social links between two groups of nodes that can be considered as a concept according to the formal concept analysis area[12]. It has been demonstrated that the set of frequent conceptual links provide a *conceptual view* of the network[40], namely a new graph structure that summarize all the knowledge acquired on the network. The conceptual view is a much reduced and semantic representation of the original network. The so-called conceptual view is a network in which a node represents a group of nodes that share common attributes and a link represents a frequent conceptual link, namely the fraction of links in the network linking these two groups.

Figure 6 depicts the conceptual view obtained from the reference network of Figure 1. We can observe that 26% of the links of the reference network connect nodes satisfying the property "X" to nodeblacks satisfying the property "Y". In the same way, 20% of the links of the reference network connect nodes satisfying the property "Y" to nodes satisfying the property "Z". We also observe that 20% of links connect nodes satisfying the property "Y".

The frequent conceptual links extraction algorithm involve two key steps: (i) a clustering phase, that built the concepts by grouping nodes on the basis of shared attributes and (ii) an evaluation process, which counts the amount of links between the concepts to evaluate the frequency of links and extract the most frequent.

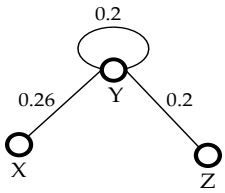

Fig. 6. Frequent conceptual links from the reference network

Although this approach is relatively new, several algorithms have been proposed for optimizing the frequent conceptual link extraction process, which proves to be a difficult task. For instance, in[39], the FCL-Min (Frequent Conceptual Link Mining) algorithm has been proposed for performing a bottom-up research while gradually reducing the search space. However, as for traditional research of frequent patterns in the data mining area, the extraction of all FCLs from a given social network may be marred by the extraction of the sub-patterns that are also frequent. Thus, in[40] Stattner and Collard present the MFCL-Min (Maximal Frequent Conceptual Link Mining) algorithm, a bottom-up method that extents the FCL-Min algorithm by focusing only on maximal frequent conceptual links (MFCL). Recently, a very promising optimization has been proposed in[41], that poses additional restrictions on candidate clusters and significantly reduces computation time needed for the extraction.

7. Conclusion

More and more works are conducted on the extraction of knowledge from social networks. As for the traditional data mining area, these works can be classified according to two broad objectives, namely the search of predictive or descriptive models. In this paper, we have presented a state of the art of current descriptive models, with a particular focus on the emerging approaches that attempt to provide models suitable to the context through the simultaneous consideration of the network structure and the attributes of nodes. Thus, we have identified and surveyed five families of methods focusing on network-based descriptive patterns: (i) link-based clustering, (ii) hybrid clustering, (iii) frequent sub-graphs discovery, (iv) network-based conceptual analysis, and (v) frequent conceptual links extraction.

A very promising and exciting path for further research on descriptive models of social networks would be to design generic models, that will be able to summarize various kind of knowledge. For instance, a first attempt has been proposed in[42] where the authors search for relevant intersections between models obtained from link-based clustering and frequent conceptual link extraction. Such an approach could allow to have, inside a single representation, several point of observations on real-world networks.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, pages 487–499, 1994.
2. A.L. Barabasi and RE Crandall. Linked: The new science of networks. *American journal of Physics*, 71:409, 2003.
3. V.D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008, 2008.
4. K. Borner, S. Sanyal, and A. Vespignani. Network science. *Blaise Cronin (Ed) Annual Review of Information Science and Technology*, 41:537–607, 2007.
5. E. Bott. *Family and social network*. The Free Press, 1957.
6. H. Cheng, X. Yan, and J. Han. Mining graph patterns. *Managing and Mining Graph Data*, pages 365–392, 2010.
7. David Combe, Christine Largeron, Elod Egyed-Zsigmond, Mathias Géry, et al. Totem: une méthode de détection de communautés adaptées aux réseaux d'information. *Extraction et gestion des connaissances (EGC'2013)*, pages 305–310, 2013.
8. Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge University Press, 2011.
9. L. Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.
10. D. Fisher and P. Langley. Approaches to conceptual clustering. In *Proceedings of the 9th international joint conference on Artificial intelligence-Volume 1*, pages 691–697. Morgan Kaufmann Publishers Inc., 1985.
11. Santo Fortunato. Community detection in graphs. *Physics Reports*, 486:75–174, 2009.
12. Bernhard Ganter, Gerd Stumme, and Rudolf Wille. Formal concept analysis, foundations and applications. *Lecture Notes in Computer Science*, 3626, 2005.
13. Bernhard Ganter and Rudolf Wille. Formal concept analysis. 45:8–13, 1996.

14. B. Gaume, E. Navarro, and H. Prade. A parallel between extended formal concept analysis and bipartite graphs analysis. *Computational Intelligence for Knowledge-Based Systems Design*, pages 270–280, 2010.
15. L Getoor and C P Diehl. Link mining: a survey. *SIGKDD Explor.*, 7:3–12, 2005.
16. Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
17. R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in organisations. *Arxiv preprint cond-mat/0211498*, 2002.
18. J. Han, H. Cheng, D. Xin, and X. Yan. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1):55–86, 2007.
19. Jun Huan, Wei Wang, Jan Prins, and Jiong Yang. Spin: mining maximal frequent subgraphs from graph databases. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 581–586. ACM, 2004.
20. A. Inokuchi, T. Washio, and H. Motoda. An apriori-based algorithm for mining frequent substructures from graph data. In *4th European Conference on Principles of Data Mining and Knowledge Discovery*, pages 13–23, 2000.
21. A. S. Klovdahl. Social networks and the spread of infectious diseases: the aids example. *Soc Sci Med*, 21(11):1203–16, 1985.
22. M. Kuramochi and G. Karypis. Frequent subgraph discovery. In *Proceedings of the 2001 IEEE International Conference on Data Mining*, pages 313–320, 2001.
23. S.O. Kuznetsov and D.I. Ignatov. Concept stability for constructing taxonomies of web-site users. *Arxiv preprint arXiv:0905.1424*, 2009.
24. B. Le Grand, M. Soto, and M.A. Aufaure. Conceptual and spatial footprints for complex systems analysis: Application to the semantic web. In *Database and Expert Systems Applications*, 2009.
25. David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.*, 58:1019–1031, 2007.
26. M. Liquiere. Some links between formal concept analysis and graph mining. *Mining graph data*, pages 227–252, 2006.
27. S. Milgram. The small world problem. *Psychology Today*, 1:61–67, 1967.
28. Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.
29. Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
30. Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.
31. Siegfried Nijssen and Joost N. Kok. The gaston tool for frequent subgraph mining. *Electr. Notes Theor. Comput. Sci.*, 127(1):77–87, 2005.
32. Claude Pasquier, Jeremy Sanhes, Frederic Flouvat, and Nazha Selmaoui-Folcher. Frequent pattern mining in attributed trees. In *PAKDD*, pages 26–37, 2013.
33. Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In *Computer and Information Sciences-ISCIS 2005*, pages 284–293. Springer, 2005.
34. Matthew J Rattigan, Marc Maier, and David Jensen. Graph clustering with network structure indices. In *Proceedings of the 24th international conference on Machine learning*, pages 783–790. ACM, 2007.
35. T.M. Riadh, B. Le Grand, M.A. Aufaure, and M. Soto. Conceptual and statistical footprints for social networks' characterization. In *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, page 8. ACM, 2009.
36. Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1706–1712, 2009.
37. V. Snasel, Z. Horak, and A. Abraham. Understanding social networks using formal concept analysis. In *Web Intelligence and Intelligent Agent Technology*, volume 3, pages 390–393, 2008.
38. V. Snasel, Z. Horak, J. Kocibova, and A. Abraham. Analyzing social networks using fca: complexity aspects. In *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, volume 3, pages 38–41. IEEE, 2009.
39. E. Stattner and M. Collard. Flmin: An approach for mining frequent links in social networks. *International Conference on Networked Digital Technologies*, 2012.
40. E. Stattner and M. Collard. Social-based conceptual links: Conceptual analysis applied to social networks. *International Conference on Advances in Social Networks Analysis and Mining*, 2012.
41. E. Stattner and M. Collard. Towards a hybrid algorithm for extracting maximal frequent conceptual links in social networks. *IEEE International Conference on Research Challenges in Information Science*, pages 1–8, 2013.
42. Erick Stattner and Martine Collard. Frequent conceptual links and link-based clustering: a comparative analysis of two clustering techniques. In *ASONAM*, pages 134–141, 2013.
43. Karsten Steinhaeuser and Nitesh V. Chawla. Identifying and evaluating community structure in complex networks. *Pattern Recogn. Lett.*, 31:413–421, 2010.
44. Y. Tian, R.A. Hankins, and J.M. Patel. Efficient aggregation for graph summarization. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 567–580. ACM, 2008.
45. Duncan J. Watts. The 'new' science of networks. *Annual Review of Sociology*, 30:243–270, 2004.
46. R. Wille. Line diagrams of hierarchical concept systems. *International classification*, 11(2):77–86, 1984.
47. Fang Wu and Bernardo A Huberman. Finding communities in linear time: a physics approach. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):331–338, 2004.
48. X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining*, 2002.
49. Jaewon Yang, Julian McAuley, and Jure Leskovec. Community detection in networks with node attributes. *arXiv preprint arXiv:1401.7267*, 2014.
50. Seok-Ho Yoon, Suk-Soon Song, and Sang-Wook Kim. Efficient link-based clustering in a large scaled blog network. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, ICUIMC '11, pages 71:1–71:5, New York, NY, USA, 2011. ACM.
51. Y. Zhou, H. Cheng, and J.X. Yu. Graph clustering based on structural/attribute similarities. *VLDB Endowment*, 2(1):718–729, 2009.