Fragile Correctness of Social Network Analysis

Mārtiņš Opmanis

Institute of Mathematics and Computer Science,University of Latvia
Rainis blvd. 29, Riga, LV1459, Latvia
martins.opmanis@lumii.lv

Abstract. Graph techniques are widely used in social network analysis. However, there are some disputable applications where results are obtained from the graphs using paths longer than one and are not simply applicable to objects in the initial domain. The author provides several examples of such usages and tries to recover roots of an incorrect application of graphs.

Keywords: Graph algorithms, social network analysis, correctness.

1 Introduction

The idea to write this paper came from quite an ordinary example of the social network depicted as the graph in Fig. 1.

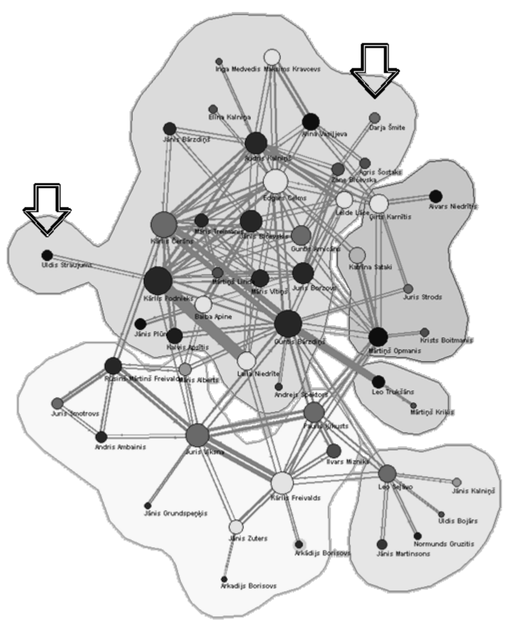

Fig. 1. Example of social network with clusters.

This network represents historical data from defence of students bachelor thesis at Faculty of Computing of University of Latvia. Supervisors and reviewers are depicted by named nodes. Arc corresponds to the particular thesis and goes from node A to node B if the first one corresponds to the supervisor and the second - to the reviewer of this thesis. There can be multiple arcs in both directions between the pair of nodes if there are several thesis with the same pair of supervisor and reviewer.

My colleagues worked on a new advanced graph visualization algorithm [1] and the given picture with clusters was an outcome of algorithm implementation. "Why these two are in the same cluster?" – I asked about two persons (depicted by arrows in Fig. 1) known by me and having no so much in common. "This is due to the algorithm. You can take a look into if you will – I'll get you source code." – responded one of colleagues[2]. At this point our conversation ended, but the received answer kept bothering me permanently for weeks and I came back to it again and again with unclear feeling that there is something completely wrong with the whole situation where developer is unable to explain outcome of this deterministic algorithm in terms of the initial task - thesis supervising and reviewing process. There is no doubt that initial picture (graph without clusters) has crystal-clear backward correspondence with real life data and at this level, there are no problems with the correctness of the data or the built graph. This is exactly the way how social network analysts process data. Then at which point this clarity is lost? Is this due to incorrect clusterization algorithm, a fault of my colleague unable to explain things in simple terms or me not catching something? Further I re-addressed the same question to main author of clusterization author. He resisted to discuss anything what led to the created graph and was not ready to discuss anything but clear graph constructs and algorithms[3].

This paper describes my investigation and connected general problems in social network analysis(SNA), and is organized as follows: in Section 2 the general process of building social networks using attributed graphs is described. In Section 3 problems with indirect ties are discussed and several examples are given. Conclusions are described in Section 4.

2 Networks as attributed graphs

"Whether studying protein interactions, sexual networks, or computer systems, the appropriate choice of representation is key to getting the correct result."[4]

Attributed graphs is one of popular forms how to model networks (author will use "network" as real world artifact). Term "graph" here is used in strong connection with graph theory and has nothing with things like infographics, charts, functions etc. Following "Today graph theory is the basis for our thinking about networks" and "The construction and structure of graphs or networks is the key to understanding the complex world around us" [5] in this paper focus will be exclusively on graphs to explain author's viewpoint. Moreover, these will be one-mode networks where all nodes and all edges are of the same kind. An example of such network is famous bridges of Königsberg mentioned in the paper of Leonhard Euler considered to be the first paper in graph theory [6]. In [7] term "network" is used in the same way as in this paper is used "attributed graph".

In the case of physical networks choice to use graphs as a source of analysis is determined by natural and simple correspondence between real life artifacts and graph constructs. Physical networks are transportation networks (pipelines, highways, waterways, airplaine routes), computer-related networks, electronic circuits and other tangible networks.

Graphs contain just two concepts – *nodes* (or vertices) which can be connected by *edges* (if undirected) or *arcs* (if directed). By adding additional attributes (like textual strings or numbers) we obtain very expressive model of an initial network – attributed graph. As well a pair of nodes may be connected by more than one edge or arc (has multiedges), etc. Precise definitions of various graph concepts can be found in [7].

For example, if road map is modeled, nodes usually denote cities and edges – direct roads connecting a particular pair of cities. Node attribute may be the name of the corresponding city, an attribute for an edge may be number and length of the corresponding road. You can use such graph-based model to investigate options how to get from one city to another – choose appropriate legs, calculate overall distance, estimate total duration of the journey. After investigation and calculations, you can simply switch back to real life and use chosen roads (corresponding to edges) for the journey. Due to a usual symmetry of roads (if you can go from A to B using the particular road, then you almost always can go also from B to A via the same road) undirected graph is an appropriate representation of road map. If in the similar way airplane routes need to be investigated, direction becomes essential – flights from airport A to airport B may have completely different schedule if compared with opposite direction and planning journey for particular day it may become essential. It must be taken into account and in the graph instead of one edge there should be two arcs with different attribute values, e.g. flight numbers, departure names, duration of flight, etc.

It is not surprising that there came idea in the same way by the graphs model interactions, relationships and ties between humans or human-based structures like companies, parties, and social groups. Excellent general overview of the history of graph usage in social network analysis is given in [8].

Despite non-physical kind of ties, some networks representing society can be considered "physical" if ties are tangible and can be proven. Kinship graph (nodes represent persons, arcs - relation "is child of"), graph of citations (nodes represent scientific publications, arcs - relation "is cited in"), World Wide Web (nodes represent webpages, arcs - relation "is linked to") are such examples.

The process of network analysis in general consists of three main steps and is schematically depicted in Fig. 2.

Three steps of SNA are:

- **"N"** – obtaining an attributed graph from the real life **n**etwork
- **"A"** – performing **a**nalysis on the created graph

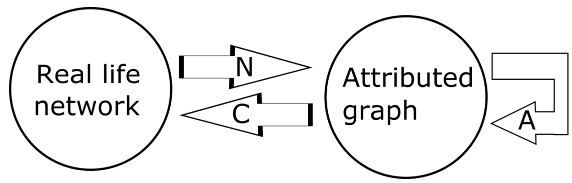

Fig. 2. Process of SNA. **"N"** – obtaining graph from the network, **"A"** – performing analysis, **"C"** – applying analysis results

– **"C"** – applying analysis results and **c**onclusions back from graph objects to real life entities

Problems with step "N" lies in untangibility of social ties. Correct representation of data has not overestimate value especially if on the basis of data analysis decisions concerning particular people or society, in general, are made [9,10]. As Edward R Tufte says, "... there are right ways and wrong ways to show data; there are displays that reveal the truth and displays that do not. And, if the matter is an important one, then getting the displays of evidence right or wrong can possibly have momentous consequences." [9].

Only in exceptional simple cases, ties will be of the same kind, e.g. "A is the scientific supervisor of B thesis". Becoming more complex, like "A send an e-mail to B" we usually need to know additional information besides the simple fact of sending particular e-mail (like having a specific subject) otherwise having the risk to obtain biased or polluted data.

Possible problems with data from social networking services are already noticed: "Unfortunately, many members of these sites try to connect with as many people as possible – whether they know them or not. This creates many false links/connections in the LinkedIn and Facebook databases. Two people might show to be connected but they really are not – one person was too embarrassed to turn down a "friend request" from a total stranger." [11]

It must be pointed out that word "attributed" here is essential. By loosing things like a name of a person, city or company, or a length of a road we obtain "bare" graph and usually lose possibility to perform step "C". Such graph as a source of information about the initial network is of limited value since several completely different parts of investigated domain may have almost the same graph topology. One of the famous examples indicating project-related interactions showing structurally similar teams from "Fortune 500" company and "Al Queda" terrorist network is depicted in [12].

If the overall goal is to describe some domain as graph without deeper analysis, you can simply omit analysis phase (step "A").

Without step "C" you can just observe obtained graph (its successful visualization may be an exceptional masterpiece!). But "The main goal of social network analysis is detecting and interpreting patterns of social ties among actors." [7] So the interpreting step "C" is essential for the whole process of SNA.

Real life artefacts (individuals, colectivities, ties, relationships, interactions etc.) are **not the same** as attributed graph concepts(nodes, edges, arcs, attributes) – the first ones are **just modelled** by laters.

In the literature, networks and graphs are usually mixed up. For example, in the introduction of [13] right after stating that "The relationships between **nodes** also have characteristics, and in network analysis we think of these as ties or links." we can read "Thus, the relationships between Bill (male, 47 years old) and Jane (female, 43 years old) may be characterized by being married, living together, co-owners of a business, and a multitude of other social ties." (p.2). It is obvious, that concept "node" belongs to graphs, Bill and Jane are representatives of real life. Social network analysts see no problem in such mixtures and talk about aspects of networks and graphs interchangeably, like "Most often, nodes **are** individuals, such as individual persons or chimpanzees." (p.2). without frustration.

This is not just fault of social network analysts. You can found "(...)where vertices **are** cities and edges **are** roads with weights *w(u,v)* denoting the length (distance) between vertices (cities) u and v(...)" in Computer Science book [14] (p.242)

Such examples of interviewing just demonstrates how natural are graph concepts fitting very well in the minds of scientists.

However, to ensure correctness of obtained results and conclusions, every transformation between real life and graphs (steps N and C) must be proven to be correct and meaningful.

Some authors hold a view that knowing graph topology is not sufficient to judge about social network properly: "...maps and metrics are mirrors, not report cards! The consultant and the client together make sense of what the maps/metrics reflect about the organization."[15]

3 Direct and indirect ties

For direct ties there is natural and strong bi-directional correspondence between graph objects and real life artifacts and raising a question about correctness seems to be ridiculous. "City A" or "node (circle, point) corresponding to city A on a map" are synonyms when you look at a simple map of a road network and cities (being very close to "pure" graph). If there is line denoting road drawn on the map between two nodes then you can be sure that also in real life such road exists and you physically can travel from A to B. As well, if a new road is built, then it will be presented in the current edition of the map. So there is almost no difference if you speak about a road on map or road in physical.

The same is true for social networks if only "observable" ties are transformed to the graph. To discover whether two persons are friends you must take a look at the corresponding attributed graph of friends, find two nodes marked by person names and they are friends only if there is edge connecting these nodes. Since such checking can be performed on any two nodes, you can ascertain that graph corresponds to the real life as far we talk just about direct ties or relationships.

Going further, for the network of roads, if there is also road from city B to city C, then you can assume that you can also get from A to C passing B. It is usually true either for humans or for vehicles. However, if instead of physical roads you investigate shemas of public transportation to get from A to C then it is possible that at B you will need to switch to different transport (from train X to train Y or even change kind of transport like from train to bus). In this case "it is possible to get from A to C" is still possible for particular person, but not for vechicle carrying passengers from A to B. This example demonstrates necessity always clearly understand modelled network.

If in this sense network is modelled correctly, we can say that relation "city X is reachable from city Y" is *transitive*, meaning that if we can reach B from A and C from B, then this imply possibility to reach C from A. Looking at the graph reachability means that nodes corresponding to particular cities *are connected* by edges either directly or indirectly (there exists a path) via other nodes.

Connectivity in graphs as well as usage of terms "walk", "trail", "path" [16, p.12] is so intrinsic that social network analysts neglect the necessity to define corresponding constructs in the initial domain and takes for granted meaningful existence of them also there. In [17] necessity to choose the right approach to characterize connectedness for indirect ties is discussed still not raising question about correctness of concept in general. In the mid-sixties term "social distance" was used to describe concept similar to "distance" in the corresponding graph[8, p.76].

In publications about social networks authors pass this topic unattended and assume that ties in the social networks have the same characteristics as in graphs and physical networks. Even if authors talk about differences between social and other networks, just numerical differences are emphasized [18] without noticing an essential differences.

However, relations may be not transitive in social networks despite connectivity in the corresponding graph! For example, from facts that A and B are friends and B and C are friends it can't be deduced that friends are also A and C! Author knows examples of several such configurations. It may be that A and C are representatives of completely different non-overlapping domains of B interests and are not familiar or state of their mutual relationship is close to "being enemies" despite knowing each other perfectly.

To analyze graph properties there is invented an overwhelming number of various graph *metrics*. Lot of them are used also for investigating social networks and according to[19], all topological metrics of distance class are based on concept "path in graph".

Problems lay in the fact that despite the structural similarity, some natural graph metrics *are not applicable* to the domain of social networks!

Observing paths and calculations of their lengths have meaning just if it possible to define how we can "move" via these paths or what is "flowing" through in the real network. Such questions never arose in physical networks - if roads are modeled, then it is possible to walk, run, ride using several roads in a row, electric current can pass several consecutive wires without questions. Physical networks (undisputable correspondence of concepts to graphs) "blindfold" SNA analysts and they overlooked this disagreement. In [13, p.3] is written about "interactions" forming "flows": "Flows may be intangibles, such as beliefs, attitudes, norms, and so on, that are passed from person to person. They can also consist of physical resources such as money or goods." Or, "Perhaps foremost among these is the idea that things often travel across the edges of a graph, moving from node to node in sequence – this could be a passenger taking a sequence of airline flights, a piece of information being passed from person to person in a social network, or a computer user or piece of software visiting a sequence of Web pages by following links." [20] "Information flows" are mentioned also in [11]: "Employees who are included in key information flows and communities of knowledge are more dedicated and have a much higher rate of retention." Obviously, interactions between particular actors (conversations, asking for advice, e-mail communication, collaborative work, etc.) are generalized to "flows" going beyond a boundary of just two involved actors. However, far from obvious is existence and quality of content of such "flows". Can we always be so sure that initial message will be passed in its original form through a long chain of actors like in the movie "Six degrees of Celebration" where it was necessary to pass the concrete message from a particular child to the president of Russia via social ties [21]? Or we will get "Chinese whispers" [22] game situation where initial message most probably is lost in the chain of transmitting people? Even assuming that people are completely honest and willing to pass information to as many relatives as possible, details usually are lost, added or transformed making almost impossible to recover the initial content of the message. Transmission of information is much more complicated and in several publications, there is described similarity in processes of spreading epidemic diseases and information[23,24]. As pointed out in [25]:"first-hand information about a disease case will lead to a much more determined reaction than information that has passed through many people before arriving at a given individual." Surprisingly, similar doubts author can find only in the famous papers describing two known **real** experiments with the usage of social ties [26,27]. These experiments have shown that there is extremally high dropout rate – the number of completed chains is just 27,5% and 19%. Judith S. Kleinfeld had found evidence that in other S.Milgrams experiments the number of completed chains was even lower (9%) and this number highly depends on such real-life attributes as race and social class[28]. On just a few experiments with dramatically low success rate lot of theory is built without further successful attempts to ground obtained results in real life.

Also back in 1967 S.Milgram noticed difference between "distance" in the real world and graph: "Almost anyone in the United States is but a few removes from the President, or from Nelson Rockefeller, but this is true only in terms os a particular mathematical viewpoint and does not, in any practical sense, integrate our lives with that of Nelson Rockefeller." [26]

Assuming that social networks can be modeled in the same way as in the physical networks is the root cause of observed problems.

One "appealing" example used in SNA is movie actor collaboration network which is built using data from the Internet Movie Database (IMDb) [29,30]. This undirected graph is built modelling actors as nodes and a particular edge connects two nodes if corresponding actors performed in the same movie. The famous parlor game "Six Degrees of Kevin Bacon" [31] is based on these data. Let's see what happens, if we try to get anything "real" from this network and apply distances to real persons.

If paths longer than 1 between actors in the network are investigated, it may lead to unacceptable consequences when transformed back to initial data – real persons. For example, famous actor Sir Thomas Sean Connery in 1957 performed in the movie "Hell Drivers" together with Wilfrid Lawson and in 1999 in the movie "Entrapment" together with Catherine Zeta-Jones [32]. Since W.Lawson and C.Zeta-Jones never performed in the same movie, distance in the one-mode network between W.Lawson and C.Zeta-Jones by definition is 2. However, in the real life W.Lawson passed away three years before C.Zeta-Jones was born (1966 and 1969 respectively), so there was no possibility for these two persons met in any way. If network shows some "transfer" of intangible things like memories, jokes, attitudes, ... then it is obvious that network should be directed and something (what?) can be carried only from W.Lawson to C.Zeta-Jones but not vice-versa! If for the moment we forget about doubts about quality of these ties and we agree that there is "something passed via this link", then reasonable is question "in **which** direction flow is going?". Since network of actor collaboration is modeled by undirected edges, we should assume that "flow" can go in any direction and each edge may be replaced by two opposite arcs. This is acceptable for direct ties - each actor may get "something" from all other actors taking part in the particular movie (assuming that this is real collaborative *work* and not just appearing together in the movie credits). However, assuming that flows can go beyond direct interaction, we come back to "flow of something" from C.Zeta-Jones to W.Lawson what is impossible.

Another popular example is joint publications [30]. Without arguing against fact that each simple fact (a collaboration between coauthors of particular publication) constituting the basis of the built network is correct – each node correspond to particular author and edge between two nodes denotes mutual publication and, most probably, also real collaborative work. Several joint publications may be represented by separate edges or by one weighted edge where weight is a number of joint publications. If analysts investigate particular author, all is correct just if they do not step over the border of distance 1 where ends collected data. Investigating things beyond this (say at the distance 2 from a particular node) mirrored back to real people needs additional explanation. The special case of collaboration network is attributed graph where "distance" from the famous mathematician Paul Erdős (1913 - 1996)[33] is investigated[20].

"Most mathematicians turn out to have rather small Erdős numbers, being typically two to five steps from Erdős. (...) The very existence of the Erdős number demonstrates that the scientific community forms a highly interconnected network in which all scientists are linked to each other through the papers they have written." [5]

But what **exactly** means "are linked through the papers" for distances greater than 1, i.e. for persons not being co-authors?

Obviously, it is assumed that real mathematicians (or scientists in general) have Erdős numbers like names and surenames - this "property" from the graph is mirrored back to real life. You can read "The point is that most mathematicians have Erdős numbers of at most 4 or 5, and – extending the collaboration graph to include co-authorship across all the sciences – most scientists in other fields have Erdős numbers that are comparable or only slightly larger; Albert Einstein's is 2, Enrico Fermi's is 3, Noam Chomsky's and Linus Pauling's are each 4, Francis Crick's and James Watson's are 5 and 6 respectively."[20]

Having lower Erdős number means producing high-quality publications? Is it enough to announce your Erdős number as a proof of quality and you will pass reviewing procedure to get published? Rather not. Similarly to problems with the network of movie actors, since the death of P.Erdős in 1996 today can not exist any tangible ties with him. If lower Erdős number implies higher scientific level, then it would be equivalent to claiming that each next generation of scientists publishing their papers is of lower scientific level if compared with the previous one (since death of P.Erdős it is impossible to get Erdős number higher than 2, after passing away of all Erdős co-authors there will be impossible to get values higher than 3, etc.). Interesting feature was noticed by my colleagues and is justification that Erdős number can not be measure of "quality" of a particular scientist - it is possible that your Erdős number is decreased (and with an appropriate assumption scientific "quality" increased) by doing absolutely nothing - it is enough if some scientist on your "social path" decrease Erdős number by publishing paper with co-author having Erdős number less than least Erdős number among all co-authors before this publication [34]. As well assigning numbers starting from P.Erdős indirectly shows that imaginary "flow" is going from P.Erdős and, if reflected back, shows that instead of true collaboration we get some "advisory flow" from co-authors with lower Erdős numbers to others. Author disagree that such kind of collaboration and spreading scientific ideas is a correct analogy for all joint publications.

One more problem is that on the basis of collaboration network there is the attempt to decide something about disciplinarity of publications [35]. Like, if there is a clique, then it can be decided that all authors from a clique are interested in and write about a similar subject. However, it is not always a case – as an example of the close scientific circle author can name himself and two persons having three pairwise connected publications [36,37,38] without any close relation of the particular topic to the third person.

It must be pointed out that there can be also undisputably tangible ties like biological kinship and in a sense of investigating spreading of genetically grounded diseases, this network is correct from the viewpoint of transitivity. Also, network of citations where nodes are publications and arcs denote referencing this publication in an another one is tangible. For paths longer than one relation “is influenced by” is completely applicable. Another example of the physical structure is the network of Wikipedia articles connected with links and “The Wikipedia Game” can be played either in the corresponding graph or using real pages and hyperlinks [39]. It should be pointed out, that in all mentioned cases ties or relationships are directed and should be modeled by arcs instead of nodes.

One of the popular ways to get information about networks in the concentrated form is to seek for communities what in the world of graphs means usage of clustering algorithms and obtaining *clusters*. General idea is that related (closely connected) objects should be included in the same cluster and distinct (weakly connected) objects – in separate clusters and there are much more edges having endpoints from the same cluster if compared with edges with endpoints from the distinct clusters. Clusters may be non-overlapping (each node belongs to at most one cluster) or overlapping (a particular node may belong to several clusters at the same time). Not so clear are criteria for division nodes in clusters. Besides simple and crystal-clear cases, like “all members of a particular clique should belong to the same cluster”, there still is a space for several distinct approaches.

Clear and unambiguous examples (like finding cliques - a group of actors having all pairwise ties) lack interest of investigators because all relationships directly follow from the initial data without the need to perform sophisticated analysis and in real social networks large cliques can be found quite rare. In the non-trivial cases are used metrics which are hard to comprehend and validate.

In later cases algorithms usually use the already discussed concept of “distance” in a network and is based on already known “paths”. In cases when paths have no reasonable meaning in the original network also task to find clusters is meaningless.

4 Conclusions

Graphs are very powerful tool for the social network analysis. However, not all cases of usage presented in publications can be admited as completely correct.

To conclude, the answer to the questions stated at the very beginning of this paper is that my colleague just followed the mainstream SNA process without noticing that some results have no sense in the original domain due to incorrect backward transformation from purely mathematical structures to initial domain.

I hope that my findings would help social network analysts to look more critically to the networks they construct as well to the quality of conclusions about networks based on completed analysis.

Acknowledgements

This work was supported by European Regional Development Fund project ????????????????????????.

The author thank Paulis Ķikusts for valuable comments.

References

1. Vihrovs, J., Prūsis, K., Freivalds, K., Ručevskis, P., Krebs, V.: An inverse distance-based potential field function for overlapping point set visualization. In: Proceedings of the 5th International Conference on Information Visualization Theory and Applications (VISIGRAPP 2014). (2014) 29–38
2. Ručevskis, P.: Personal conversation (2013)
3. Ķikusts, P.: Personal conversation (2013)
4. Butts, C.T.: Revisiting the foundations of network analysis. Science **325**(5939) (2009) 414–416
5. Barabasi, A., Frangos, J.: Linked: The New Science Of Networks Science Of Networks. Basic Books (2014)
6. Hopkins, B., Wilson, R.J.: The Truth about Königsberg. The Colledge Mathematics Journal **35**(3) (5 2004) 198–207
7. de Nooy, W., Mrvar, A., Bategelj, V.: Exploratory Social Network Analysis with Pajek. 2 edn. Cambridge University Press (2012)
8. Scott, J., ed.: Social Networks: Critical Concepts in Sociology. Volume 1. Routledge (2002)
9. Tufte, E.R.: Visual Explanations: Images and Quantities, Evidence and Narrative. Graphics Press, Cheshire, CT, USA (1997)
10. Robinson, W., Boisjoly, R., Hoeker, D., Young, S.: Representation and Misrepresentation: Tufte and the Morton Thiokol Engineers on the Challenger. Science and Engineering Ethics **8**(1) (2002) 59–81
11. Krebs, V.E.: Social capital: the key to success for the 21st century organization. IHRIM **XII**(5) (2008) 40
12. Krebs, V.: Two project teams. `http://orgnet.com/WhichOne.pdf` (2002)
13. Borgatti, S.P., Everett, M.G., Johnson, J.C.: Analyzing Social Networks. SAGE publications Ltd. (2013)
14. Reus, B.: Limits of Computation: From a Programming Perspective. Springer International Publishing Switzerland (2016)
15. Krebs, V.: Social network analysis: An introduction by orgnet, llc. `http://www.orgnet.com/sna.html` (2002)
16. Bondy, J.A., Murty, U.S.R.: Graph Theory With Applications. Elsevier Science Publishing Co., Inc. (1976)
17. Peay, E.R.: Connectedness in a General Model for Valued Networks. Social Networks (2) (1980) 385–410
18. Newman, M.E.J., Park, J.: Why social networks are different from other types of networks. Physics Review (2003) E 68:036122.
19. Hernández, J.M., Mieghem, P.V.: Classification of graph metrics. Technical report, Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, 2628 CD Delft (November 2011)
20. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press (2010)
21. Bekmambetov, T., Chevazhevskiy, Y., Jonynas, I., Kiselev, D., Voytinskiy, A.: Movie "Six Degrees of Celebration" (original title – "Yolki"). `http://www.imdb.com/title/tt1782568/` (2010)
22. Blackmore, S., Dawkins, R.: The Meme Machine. New ed edn. Oxford University Press (2000)
23. Goffman, W., Newill, V.A.: Communication and Epidemic Processes. Proceedings of the Royal Society of London Series A **298** (May 1967) 316–334

24. Goffman, W.: A mathematical method for analyzing the growth of a scientific discipline. J. ACM **18**(2) (April 1971) 173–185
25. Funk, S., Gilad, E., Watkins, C., Jansen, V.A.A.: The spread of awareness and its impact on epidemic outbreaks. Proceedings of the National Academy of Sciences **106**(16) (2009) 6872–6877
26. Milgram, S.: The Small World Problem. Psychology Today **2** (1967) 60–67
27. Travers, J., Milgram, S.: An Experimental Study of the Small World Problem. Sociometry **32**(4) (December 1969) 425–443
28. Kleinfeld, J.S.: The Small World Problem. Society **39**(2) (January 2002) 61–66
29. Needham, C.: Internet movie database. `http://www.imdb.com` (1998)
30. Borenstein, E.: University of Washington course GS559: Introduction to Statistical and Computational Genomics (Winter 2016), Slides of lecture 15: Biological networks and Dijkstra's algorithm (2016) `http://elbo.gs.washington.edu/courses/GS_559_16_wi/slides/15A-Networks_Dijkstra.pdf`.
31. Fass, C., Turtle, B., Ginelli, M.: Six Degrees of Kevin Bacon. Plume (1996)
32. Connery, S.: Filmography. `http://www.seanconnery.com/filmography/` (2016)
33. Erdős, P.: Wikipedia. `https://en.wikipedia.org/wiki/Paul_Erd%C5%91s` (2016)
34. Ručevskis, P., Podnieks, K., Kozlovičs, S., Grasmanis, M., Celms, E.: Personal conversation (2016)
35. Fortunato, S.: Community detection in graphs. Physics Reports **486** (February 2010) 75–174
36. Viksna, J., Celms, E., Opmanis, M., Podnieks, K., Rucevskis, P., Zarins, A., Barrett, A., Neogi, S.G., Krestyaninova, M., McCarthy, M.I., Brazma, A., Sarkans, U.: Passim – an open source software system for managing information in biomedical studies. BMC Bioinformatics **8**(1) (2007) 1–7
37. Opmanis, M., Čerāns, K.: Multilevel data repository for ontological and meta-modeling. In: Databases and Information Systems VI-Selected Papers from the Ninth International Baltic Conference, DB&IS. (2010)
38. Čerāns, K., Vīksna, J.: Deciding reachability for planar multi-polynomial systems. In Alur, R., Henzinger, Thomas A.and Sontag, E.D., eds.: Hybrid Systems III: Verification and Control. Springer Berlin Heidelberg, Berlin, Heidelberg (1996) 389–400
39. Clemesha, A.: The wikipedia game `http://thewikigame.com/`.