Alguliev R.M.[1], Aliguliyev R.M.[2], Ganjaliyev F.S.[3]
Institute of Information Technology of ANAS, Baku, Azerbaijan
[1]rasim@science.az; [2]a.ramiz@science.az; [3]fadaig@yahoo.com

SOCIAL NETWORK ANALYSIS ON THE WEB: STATE OF THE ART, PROBLEMS AND FUTURE DIRECTIONS

Social networks have attracted much attention recently. Different studies have been conducted to automatically extract social networks among various kinds of entities from the Web. Social network analysis (SNA) finds its application in many current business areas. Although much work has been done for extraction and analysis of social networks on the Web, still there are open problems for which solutions are required. In this paper initially we give a brief history of social network analysis. Then, we describe some of the mostly used data sources for social network data on the Web. Afterwards, we give an overview of some of the existing methods and algorithms for extracting social networks from the Web. Finally, we conclude with what we shall be addressing in our research.

***Key words:** web-based social networks; data source; social network analysis; actors*

1. Introduction

A social network is a set of relationships. More formally, a social network contains a set of objects (in mathematical terms, nodes) and mapping or description of relations between the objects or nodes. The simplest social network contains two objects, 1 and 2, and one relationship that links them. Nodes 1 and 2, for example, might be people, and the relationship that links them might be “are standing in the same room”. There are also directional relationships such as 1 likes 2. In this simple network, the relationship could be symmetrical or non-directional: 1 and 2 like each other, or their liking is mutual. There need not be just one relationship mapped between nodes 1 and 2. For example, 1 and 2 might be in the same room and might like each other. When there is more than one relationship, this is called a *multiplex* relationship.

In social network theory network graphs are often referred to as *sociograms*. In some cases it would be more convenient to represent a social network with the aid of a matrix. The most common form of matrix in social network analysis is a very simple one composed of as many rows and columns as there are actors in our data set, and where the elements represent the ties between the actors. The simplest and most common matrix is binary. That is, if a tie is present, one is entered in a cell; if there is no tie, a zero is entered. This kind of matrix is the starting point for almost all network analysis, and is called an ”adjacency matrix”, because it represents who is next to, or adjacent to whom in the ”social space” mapped by the relations, that we have measured [1].

The appearance of web put a series of interesting challenges to the researches of social networks. First of the entire web influenced the traditional way of thinking about social networks. Since SNA usually was conducted on small group of nodes, it appeared to be not so easy to apply the same methods on the Web. In most cases, it was nearly impossible to analyze a network of millions of people taking into account that the proper analysis requires construction most of the network. Another hard task was to gather information about a large group of people. Moreover, there can be and actually are people who are actors of the network but do not need any generated data which requires specific considerations in this case. These types of networks are much more difficult to study [2].

2. Social network analysis

The origin of contemporary SNA can be traced back to the work of Stanley Milgram. In his famous 1967 experiment, Milgram conducted a test to understand how people are connected to others by asking random people to forward a package to any of their acquaintances, who they thought might be able to reach the specific target individual [3]. In his research, Milgram found that most people were connected by six acquaintances. This research led to the famous phrase "six degrees of separation," which is still widely used in popular culture.

Another important step in the development of SNA was the work of Mark Granovetter on network structures. In his widely-cited 1973 article "The Strength of Weak Ties," Granovetter argues that "weak ties" – your relationships with acquaintances – are more important than "strong ties" – your relationships with family and close friends – when trying to find employment [4].

The primary concern of SNA is not the attributes of actors but the nature of relations. This means that except for measuring existing relation between pairs of nodes, it is important to understand the effect and meaning of a tie.

When studying properties of networks as a whole, researchers can look at such things as the proportion of dyads connected to one another (*density*), the average path length necessary to connect pairs of nodes, the average tie strength, or the extent to which the network is composed of similar nodes (*homogeneity*) or nodes with particular characteristics, such as the proportion of network members who are women (*composition*).

In addition, networks can be studied in the ways that they can be divided into sub-graphs. For example, networks may consist of multiple *components*: sets of nodes that are tied directly or indirectly to one another but are not tied directly to nodes in other components. They may also include *cliques*, in which every node is tied directly to every other node [5].

3. Data source for SNA on the Web

There exiss a variety of data sources on the Web from which social networks can be extracted and further analysis can be performed. *Forums* represent one of the sources of data for extracting social networks on the Web. Some people consider them as message boards or bulletin boards. Others name it threaded discussions, discussion boards or discussion groups or even conferences. The simplest definition of a forum can be a place where people have the chance to communicate and reply to each other's threads. A forum member posts a message usually on a specific topic which is accessible to everyone else in the forum. After the post is read the reader has the option to leave a reply to the message which is in turn also visible to the entire forum. In this way a discussion is created. It does not require all the users to be online constantly. They may leave a message, go offline and later come online to view replies if any.

Blogs are websites in which the individual writes, publishes and distributes his/her own opinions via the internet on any subjects. To create a blog one does not need any technical knowledge. One usually adds pages and articles through a Windows type of interface in such a way there is no need for a web designer to update it. The blog visitors are able leave replies to posts created by the blog leader. Search engines obviously can access blogs as any other sites. It is also possible that not only one person can add posts but you can allow multiple people to create posts i.e. to have multiple authors to exist if they interested in the topic.

Social networking sites and social media sites have had both positive and negative impact. Many users spend excessive amounts of time creating and/or viewing self- aggrandizing or totally trivial updates on tiny details of their and their so-called online friends' activities. Many of them lose sense of what is appropriate, and end up losing employment or college entrance opportunities because of their having posted inappropriate remarks or photos on social networking sites or social media sites.

Social Web sites provide various facilities for members to use to communicate with their online connections, that is, friends and other members. These include email, instant messaging, text messaging, and public and private bulletin boards, and even Internet phone services. Such sites as MySpace and Facebook allow their members to use the messaging and phone call facilities of Internet phone services, such as Skype.

Social Web sites, as we know them, today have only a 5–6 year history. The growth of some of the sites, in terms of the number of users, the level of daily traffic, and the amount of UCCs (User Created Content) stored, has been absolutely incredible. The general sense is that even the founders of those sites did not envision such growth. As social Web sites have evolved, various unforeseen issues and challenges have surfaced [6].

Virtual teams are one of interesting areas on the Web for social network researchers. These are groups of people who usually gather together over time but basically work at some distance from each other. Often they may never meet each other. Using different data sources, one of the network researchers' objectives is to develop method to detect such groups of people.

Other sources of data for network analysis on the Web include the network of academic researchers which co-authored papers together. Also there is so-called co-citation network which is a network of researchers who have cited other papers in their own articles. Similarly, rich set of data is available for analysis in collaboration networks, funding networks, sporting networks and etc.

4. The state of the art of SNA on the Web

Basically there are two methods for retrieving social networks: interview methods and automatic extraction methods. Interview methods consist of asking the actors to answer particular questions, like "Who do you go to when you encounter a problem in your daily work?". The primary disadvantage of this method is obviously the incorrectness of information. We can never guarantee that all respondents answered what the truth really is. One may simply incorrectly specify a name; another one can do it intentionally. Automatic extraction is widely applied on the Web currently.

Social networks on the web have been extracted by retrieving of relationships between entities all automatically derived from multilingual news. In [7], Pouliquen et al. have built a system which extracts social networks from news articles. Daily Europe Media Monitor daily collects 50,000 newspaper articles in 40 different languages from 1,500 websites around the world. Each received article is tagged with entities that it mentions. Two types or relations are extracted from the articles: co-occurrence relationship and qualified relationship. Social networks also have been extracted from log files of online shared workspaces [8]. Another method is used to extract biographical information of historical persons from multiple unstructured sources on the Web [9]. Extraction of social networks has also been conducted via Internet and Networked Sensing [10].

Kautz et al. [11] designed a social network extraction system from the Web, named *Referral Web*. As an evidence of a relationship between two persons, the system uses co-occurrence of names in almost any resource available on the Web which includes links on homepages, list of co-authors in technical papers and citations of papers, any communication in news archives between persons and organization groups.

Another system for extracting social networks from the Web, named Flink, was developed by Mika [12]. The system extracts social network using Web pages, publications, e-mail messages and FOAF (Friend Of A Friend) files. In order to extract a relationship between entities the system uses a general search engine by means of which it takes hit counts for separate actors and for their co-occurrence. Since the system aims at Semantic Web the keyword, "Semantic Web OR Ontology" is used when submitting queries to the search engine. This is done in order to avoid ambiguity of named entities.

A system which extracts social network from user's inbox was developed by Culotta et al. [13]. This system identifies unique people in e-mail pages, finds their homepages, assigns keywords to each contact in the network, and fills the fields of a contact address book. It also establishes links between the owner's of the Web pages and persons found on that page.

Matsuo et al. [14] developed a system Polyphonet which is used for extracting and analyzing a social network of academic figures. The system allows performing searches of different types: by name, by keyword, affiliation and research field. It also lists the researchers related to the searched person. The system uses the overlap coefficient to measure the strength between entities. Several algorithms are used to extract a social network of a researcher in Polyphonet.

Tang et al. [15] demonstrates a method for extracting an academic social network. The approach is to treat a researcher's network extraction as a procedure containing three steps: identification of relevant Web pages, profile information extraction and integrating publications from different sources. The relevant page identification is carried out by searching pages containing the researcher's name through a search engine. Publications are not retrieved directly from Web but from online data sources. The disambiguation problem is also addressed by regarding it as a problem of assigning some *n* publications to their real researchers given a person name *a* and publications containing the name *a*.

Based on the method of spanning trees which reduces the clustering problem to a graph partitioning problem, the work done in [16] detects groups between researchers in Computing Science in Brazil. When extracting social network of researchers from the Web it uses four types of relationships: advisoring work, Project Participation, technical production and publication co-authorship. Edge strength of two entities is assigned by first dividing the number of common relationships between the two entities by the sum of publications of the first and researchers.

In [17], the author demonstrates how a social network can be extracted from email communication of users. If two users exchanged more than some *N* number of emails, then ad edge is invented between them. In order to assign ranks to users two statistics are measured. First is based on the hypothesis, that if two users communicate more, then they should exchange more emails. So the first statistics is simply the number of emails a user has sent or received. The second statistics is the average response time. This, in fact, is the time a user waits when another one responds to his or her email from the moment the first one has sent the email. The reply is considered to be a response if the sender receives a message within three business days. In this way, a user is assigned the statistics of the average response time based on all responses from other users. After the social network has been built, all cliques are obtained by means of the method described in [18]. Users ranked basing on the degree centrality, betweenness centrality and also on the number of cliques a user is contained within, the size of the clique and the weights of the clique edges.

One of the most popular searches on the Web is searching people related to some other person. For example, people need to find other people who have published more papers in a specific topic or to find the most famous actor in a certain area. Author of [19] proposes an approach to resolve this problem. Two issues are presented as being primary in searching people sharing similar interests: person representation and matching person. The approach of presenting a person on the Web is to use the person's Web site since this usually contains information about the person's background and interests. In the first work, four algorithms are presented for finding people with similar interests. In the first algorithm the content and links on all pages on the person's web site are used and integrated in order to represent a person. In the second algorithm only the textual content of all pages on the web site are explored. The third algorithm uses only the link structure of all pages on the web site. Finally, the fourth algorithm uses both the textual content and link structure of only the main page of the person's web site. Individually, the content-based approach and link-based approaches have some disadvantages like word mismatches, few links and etc. For this reason the author hypothesizes, that integrating these both methods will outperform other three algorithms.

On the Web extracting social networks from FOAF documents is a new research topic. Some studies have extracted networks from FOAF documents [20], [21]. For example, Aleman-Meza et al. proposes a method which combines two social kinds of relationships: FOAF network and co-authorship network. They integrate two social networks to determine the degree of conflict of interest among scientific researchers [22].

One of the most important questions in social network analysis research is to define the most influential actor, i.e. to find the node which has the most influence in the network. In network theory this is referred to as *centrality*. The most general measures of centrality are (as defined by Freeman) *degree* centrality, *closeness* centrality and *betweenness* centrality. Centrality tells us about key actors in the network, i.e. which actors have the highest ranks. Ranking is one of the most important functions which need to be computed when conducting social network analysis. Ranking shows which actors in the network are most important in the sense that which are the most influential, which take decisions and etc. When determining centrality of a node in a network, all neighbour nodes of the node are taken into consideration, as well. Centrality is highly dependent on the strength of the nodes of the network. In order to calculate edge strength, several functions exist as Jaccard, overlap, Dice and etc. These similarity measures accept two sets of data and return the estimated strength between them. This, in its turn means that the similarity measure chosen to calculate network edge strengths may affect the results of ranking, i.e. as a result of network analysis actors may be assigned various ranks depending on the chosen similarity coefficient.

In [23], it is demonstrated how ranking results of nodes can vary depending on which similarity coefficient one chooses to calculate edge weights in the network. In this work three social networks of arbitrary five participants of the WWW2006 conference are built. Then, nodes are ranked using the Topic-Centric algorithm. Afterwards, for each of the networks a new network is derived using the concept of resistance distance. In these three networks the nodes are ranked again using the same algorithm. The analysis showed that there might be a difference in the ranking process i.e. actors with some ranks may receive higher or lower ranks depending on the chosen similarity function.

Academic researchers may have relations of different kinds: they may have one or more co-authored papers, may participate in the same conference, may be members of the same scientific centre or might have taken part in the same project and etc. Extraction of social network of academic researchers is also addressed in some works [14], [15], [24]. The major issue with the work done on extracting social network of academic researchers on the Web is that they do not embrace the entire academic network. One of the reasons of this is that semantic based information is not enough to extract complete information about academic networks. Also, data generated by users is not always consistent.

A new method for building a social network of academic researchers is presented in [24]. The idea behind this method is as follows. It is assumed that a set of named entities for which a social network needs to be built is given. For each of the supposed relations a separate single-relational network is built. After this, all the received networks are integrated into a multi-network as demonstrated in the work. It is hypothesized that the obtained network will provide better results when conducting analysis of the corresponding integrated network.

Another direction in the social network research on the Web is community mining. Community mining is one of primary research areas in SNA. One may define community as a set of objects sharing similar properties. In graph theory concepts community mining can be defined as identification of sub-graphs, i.e. identifying sub-graphs which have higher densities compared to the whole network density. Some works on the Web address community mining problem, as well [25], [26], [27].

Although not little work done on detecting social networks on the Web, most of them assume that there is only one relationship between entities in the graph. In fact, in most communities, the nodes are connected by more than one relation. Moreover, each relation has a particular role in a particular task. When supposing that in a given community only one relation exists between people, much of valuable information can be lost. Besides, many of these algorithms are not suitable for large-scale networks and concentrate on small networks.

5. Problems and future directions

Most methods and algorithms which exist for SNA on the Web assume that there is only one relationship between actors. In real, more than one relation exists in social networks. Moreover, most methods proposed are well for specific sets of data like news articles, researchers and etc. and don't demonstrate expected results for other kind of data. Besides, most algorithms are based on co-occurrence based approach. The idea behind this approach is that if two entities on the Web are mentioned on one and the same page then they have some relations in common.

Whichever method is used for extracting social network, it is restricted by the available data. For example, survey methods may give unexpected network analysis results for a simple reason that we don't have guarantee whether the interviewee ignored something or not. The interviewee may not be willing to provide some private information. Drawback of automatic extraction methods is, for example, for famous persons it may give good results but may not work well for people not mentioned much on the Web. Co-occurrence based approach, for example, in this case, may not work well. Other integration methods might be necessary. Similarly, on social networking sites users have the option to hide a part of information from other users and make another part of information available to public.

Taking into account the above mentioned questions and existing methods for extracting and analysis of social networks on the Web, in our research, we aim at addressing the following tasks:

- Develop methods and algorithms for extracting social network of academic researchers from the Web.
- Analyse of the role of similarity measures in social network synthesis.
- Develop methods for detecting communities on the Web.
- Develop algorithms for SNA on the Web.

As a first task we shall try to find out the basic symptoms of network analysis when different similarity measures are chosen to find edge weights for one and the same network, i.e. having one and the same set of nodes we shall assign weights to edges of a network using particular similarity measure. Afterwards, we shall apply a different similarity coefficient to the same set of nodes and so on. Finally, we shall compare the results received and try to predict the outcome.

Secondly, we shall suggest new approaches and develop new methods for extracting social network of academic researchers from the available information on the Web. We shall try to work out methods which would extract the whole network of researchers.

Since community mining is one of interesting and primary research areas in social network analysis on the Web and still much work needs to be done in this area, we shall also address this field and try to develop methods which would have effective results in retrieving communities from the Web.

After all the above objectives are accomplished, we shall work out some methods which would allow carrying out analysis of the networks extracted. At this point, we shall concentrate at networks for extraction of which the methods and algorithms we have suggested.

6. Conclusion

Recently social networks have gained much popularity. Different studies have been conducted to automatically extract social networks among various kinds of entities from the Web. Various data source exist on the Web from which social network can be extracted and further analysis can be performed. Despite not little work done to extract social networks on the Web, they all have some disadvantages. Effect of ranking model on network analysis, extraction of a network of academic communities, and also detecting communities is also performed on the Web. Nevertheless, some problems exist in the proposed methods. During our research period, we shall try to address the problems we have specified above.

7. References

[1] M.Jamali and H.Abdolhassani, Different aspects of social network analysis, *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, Hong Kong, China, 2006, pp.66–72.

[2] J.Golbeck, Web-based social networks: a survey and future directions, *Technical Report*, 2005, 14 p.

[3] M.Gotta, Analysis of Social Networks: telling old stories in new ways, (http://mikeg.typepad.com/perceptions/2008/04/analysis-of-soc.html), 2008.

[4] M.Granovetter, The strength of weak ties, *American Journal of Sociology*, 1973, v.78, no.6, pp.1360-1380.

[5] A.Marin and B.Weldman, Forthcoming in handbook of social network analysis, SAGE, London, 2000, 240p.

[6] W.Kim, O.Jeong, and S.W.Lee, On social web sites, *Information Systems*, 2010, vol. 35, no. 2, pp. 215–236.

[7] B.Pouliquen, H.Tanev, and M.Atkinson, Extracting and learning social networks out of multilingual news, *Proceedings of the Social Networks and Application Tools Workshop (SocNet-08)*, Skalica, Slovakia, 2008, pp.13–16**.**

[8] P.Nasirifard, V.Peristeras, C.Hayes, and S.Decker, Extracting and utilizing social networks from log files of shared workspaces, *IFIP Advances in Information and Communication Technology*, 2009, vol.307, pp.643–650.

[9] G.Geleijnse, and J.Korst, Creating a dead poets society: extracting a social network of historical persons from the web, *Proceedings of the 6th International Conference of the Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference*, Busan, Korea, 2009, pp.156–168.

[10] T.Nishimura and Y.Matsuo, A method of social network extraction via internet and networked sensing, *Proceedings of the 3rd International Conference on Networked Sensing Systems*, Chicago, USA, 2006. http://citeseerx. ist.psu.edu /viewdoc/download ?doi= 10.1.1.108.3289&rep=rep1&type=pdf

[11] H. Kautz, B.Selman, and M.Shah, The hidden web, *AI Magazine*, 1997, vol. 18, no.2, pp.27–35.

[12] P.Mika, Flink: semantic web technology for extraction and analysis of social networks, *Journal of Web Semantics*, 2005, vol.3, no.2, pp.211–223.

[13] A.Culotta, R.Bekkerman, and A.McCallum, Extracting social networks and contact information from email and the web, *Proceedings of the First Conference on Email and Anti-Spam*, California, USA, 2004. http://www2.selu.edu/Academics /Faculty/ aculotta/ pubs/culotta04extracting.pdf

[14] Y.Matsuo, J.Mori, and M.Hamasaki, POLYPHONET: an advanced social network extraction system, *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, 2006, pp.262–278.

[15] J.Tang, D.Zang, and L.Yao, Social network extraction of academic researchers, *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, Omaha, USA, 2007, pp.292–301.
[16] V.Stroele, J.Oliveira, G.Zimbrao, and J.M. Souza, Mining and analysing multirelational social networks, *Proceedings of the 2009 International Conference on Computational Science and Engineering*, Vancouver, Canada, 2009, pp.711–716.
[17] R.Rowe, G.Creamer, S.Hershkop, and S.J.Stoflo, Automated social hierarchy detection through email and network analysis, *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, San Jose, USA, 2007, pp.109–117.
[18] C.Bron, J.Kerbosch, Finding all cliques of an undirected graph, *Communications of the ACM*, 1973, vol.16, no.9, pp.575–577.
[19] Q.Li and Y.B.Wu, People search: searching people sharing similar interests from the Web, *Journal of the American Society for Information Science and Technology*, 2007, vol.59, no.1, pp.111–125.
[20] T. Finin, L. Ding, and L.Zou, Social networking on the semantic Web, *The Learning Organization*, 2005, vol.12, no.5, pp.418–435.
[21] B.Aleman-Meza, M.Nagarajan, L.Ding, A.Sheth, et al, Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection, *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, Scotland, 2006, pp.407–416.
[22] J.Goldbeck, and M.Rothstein, **Linking social networks on the web with FOAF,** *Proceedings of the 23rd National Conference on Artificial Intelligence*, 2008, Chicago, USA, pp.1138–1143.
[23] R.Alguliyev, R.Alguliyev, and F.Ganjaliyev, Role of similarity measures in ranking network entities, *Intelligent Information Management*, 2010 (in press).
[24] F.Ganjaliyev, Building a heterogeneous social network of academic researchers, *Proceedings of the 3rd International Conference of "Problems of Cybernetics and Informatics"*, Baku, Azerbaijan, 6–8 September, 2010, vol.1, pp.179–182.
[25] N.Du, B.Wu, X.Pei et al., **Community detection in large-scale social networks,** *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, San Jose, USA, 2007, pp.16–25.
[26] J.Baumes, M.Goldberg, M.Magdon, and W.Wallace, Discovering Hidden Groups in Communication Networks, *Proceedings of the 2nd NSF/NIJ Symposium on Intelligence and Security Informatics*, Tucson, USA, 2004 pp.378–389.
[27] D.Cai, Z.Shao, X.He et al., Mining hidden community in heterogeneous social networks, *Proceedings of the 3rd International Workshop on Link Discovery*, 2005, Chicago, USA, pp.58–65

UOT 004.8
Əliquliyev Rasim M.[1], Alıquliyev Ramiz M.[2], Gəncəliyev Fədai S.[3]
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
[1]rasim@science.az; [2]a.ramiz@science.az; [3]fadaig@yahoo.com

Veb-də sosial şəbəkələrin analizi: müasir vəziyyət, problemlər və perspektivlər

Son zamanlar sosial şəbəkələr daha çox diqqət cəlb edir. Veb-də sosial şəbəkələrin avtomatik aşkarlanması üçün müxtəlif tədqiqatlar aparılmışdır. Hal-hazırda sosial şəbəkələrin analizinə geniş yer verilir. Veb-də sosial şəbəkələrin aşkarlanması və analizi ilə bağlı aparılmış kifayət qədər işlərə baxmayaraq, bəzi problemlər indiyə kimi həllini tapmamışdır. Məqalədə ilk növbədə sosial şəbəkələrin analizinin qısa tarixi verilmiş, bundan əlavə sosial şəbəkələrin aşkarlanması üçün geniş istifadə olunan bəzi mənbələrin təsnifatı verilmişdir. Sonda Veb-də sosial şəbəkələrin

aşkarlanması üçün mövcüd olan metod və alqoritmlərə baxılmış və həlli vacib olan məsələlər sadalanmışdır.
Açar sözlər: *Veb-də sosial şəbəkələr; verilənlər mənbəyi; sosial şəbəkələrin analizi; aktorlar*

УДК 004.8
Алгулиев Расим М.[1], Алыгулиев Рамиз М.[2], Гянджалиев Федаи С.[3]
Институт Информационных Технологий НАНА, Баку, Азербайджан
[1]rasim@science.az; [2]a.ramiz@science.az; [3]fadaig@yahoo.com
Анализ социальных сетей на Веб: современное состояние, проблемы и перспективы
В последнее время социальные сети привлекают все большее внимание. Разные исследования были проведены для автоматического извлечения социальных сетей между разными видами ресурсов на Веб. Анализ социальных сетей находит широкое применение в настоящее время. Несмотря на то, что проделано достаточно работы для извлечения и анализа социальных сетей на Веб, некоторые проблемы остаются все еще открытыми. В данной статье сначала дается краткая история анализа социальных сетей. Затем описываются некоторые из наиболее используемых источников данных для извлечения социальных сетей на Веб. Далее рассматриваются некоторые из существующих методов и алгоритмов для извлечения социальных сетей на Веб. В заключении перечисляются актуальные задачи в данной области.
Ключевые слова: *социальные сети на Веб; источник данных; анализ социальных сетей; акторы*