Annales Univ. Sci. Budapest., Sect. Comp. 43 (2014) 57–68

COMMUNITY SHELL'S EFFECT ON THE DISINTEGRATION OF SOCIAL NETWORKS

Imre Szücs (Budapest, Hungary)

Attila Kiss (Budapest, Hungary)

Dedicated to András Benczúr on the occasion of his 70th birthday

Communicated by Péter Racskó

(Received June 1, 2014; accepted July 1, 2014)

Abstract. The stability of social networks is an important question from several points of view. It is important to find the most attackable parts of networks or to find the most important customers of a commercial company. Several methods exist for ranking the nodes of a network based on different aspect of importance and also for detecting communities of a network. In this paper we examine the relationship between communities and centrality measurements from the network stability point of view. We have found that the clusters behave as a shell around its members. The dynamic of the disintegration of the network strongly depends on the used centrality measure and the usage of communities.

1. Introduction

Stability analysis, importance of nodes and community detection are important questions of applied network theory. Several applications of social network

Key words and phrases: social network, community detection, stability analysis, graph centrality

2010 Mathematics Subject Classification: 91D30, 91C20

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013). This work was completed with the support of the Hungarian and Vietnamese TET (grant agreement no. TET 10-1-2011-0645).

analysis use these concepts, like recommendation systems, churn models, up- and cross-sell models, where it is important to know what is the most likely to happen with the network in case an element of the network disappears. Beside commercial applications the network security and economics domain is also mentionable field where the research of network stability is in the focus of interest.

Many studies focus on the examination of social networks as complex dynamical systems [1]. In [2] the differences between social and web graphs were examined and showed its effect on the determination of the importance of nodes. The result of [3] confirmed the power-law, small-world, and scale-free properties of on-line social networks and discussed the implications of the structural properties for the design of social network based systems. In [4] the micro-level network properties were studied and the centrality measures were suggested for using as indicators for impact analysis.

2. Data and methodology

For analyzing the disintegration dynamic of networks 3 different datasets and 2 vertex ranking methods were used. To measure the effect the number of communities were used based on the walktrap community detection method.

2.1. Datasets

2.1.1. Amazon product co-purchasing network

The Amazon product co-purchasing network was collected by crawling Amazon website. In case a product i is frequently co-purchased with product j, the graph contains an undirected edge from i to j [5].

Table 1 contains the main statistics of the Amazon product co-purchasing network.

2.1.2. Condense Matter collaboration network

The Condense Matter collaboration network covers scientific collaborations between authors of papers submitted to Condense Matter category. In case an author i co-authored a paper with author j, the graph contains a undirected edge from i to j. The data covers papers in the period from January 1993 to April 2003 (124 months) [6].

Nodes	334863
Edges	925872
Nodes in largest WCC	334863
Edges in largest WCC	925872
Nodes in largest SCC	334863
Edges in largest SCC	925872
Average clustering coefficient	0.3967
Number of triangles	667129
Fraction of closed triangles	0.07925
Diameter (longest shortest path)	44
90-percentile effective diameter	15

Table 1. Statistics of dataset: Amazon product co-purchasing network

Table 2 contains the main statistics of the Condense Matter collaboration network.

Nodes	23133
Edges	93497
Nodes in largest WCC	21363
Edges in largest WCC	91342
Nodes in largest SCC	21363
Edges in largest SCC	91342
Average clustering coefficient	0.6334
Number of triangles	173361
Fraction of closed triangles	0.107
Diameter (longest shortest path)	14
90-percentile effective diameter	6.5

Table 2. Statistics of dataset: Condense Matter collaboration network

2.1.3. High-energy physics citation network

The High-energy physics citation network covers all the citations within a dataset of 34,546 papers with 421,578 edges. If a paper i cites paper j, the graph contains a directed edge from i to j. The data covers papers in the period from January 1993 to April 2003 (124 months) [7], [8].

Table 3 contains the main statistics of the High-energy physics citation network.

Nodes	34546
Edges	421578
Nodes in largest WCC	34401
Edges in largest WCC	421485
Nodes in largest SCC	12711
Edges in largest SCC	139981
Average clustering coefficient	0.2848
Number of triangles	1276868
Fraction of closed triangles	0.05377
Diameter (longest shortest path)	12
90-percentile effective diameter	5

Table 3. Statistics of dataset: High-energy physics citation network

2.2. Methodology

2.2.1. Community detection

For examining the actual number of communities the Igraph [9] implementation of the Walktrap algorithm [10] was used. The idea behind the algorithm, is that is more likely to stay inside a community than to move outside from it when walking over the nodes. To map the network short (3-4-5 step) random walks are used.

2.2.2. Centrality measure

For centrality measure the Igraph [9] implementation of degree (both in- and out-degree) and Pagerank algorithm were used. The degree of a vertex is its most basic structural property, the number of its adjacent edges. The Pagerank algorithm is described in details in [11].

2.2.3. Measure of disintegration

The main goal was to examine the dynamic of the disintegration, when iteratively eliminating nodes from the network, based on different strategies. In each strategy 100 iterations were executed and the 10 most important nodes were removed in each iteration, while the structure of the graph was characterized by the actual number communities.

In one hand three ranking methodologies were used in our research. As a benchmark strategy a random elimination was used firstly when 10 nodes were removed randomly in each iteration. Then a degree based strategy was used, where the 10 nodes with the highest degree were removed during the iterations.

In the Pagerank based strategy the nodes to remove were selected based on the Pagerank values of the nodes.

On the other hand three structural strategies were used as well, where the set of nodes used in the calculations were defined differently. In the Graph-level strategy the importances of the nodes were calculated from the whole graph. In case of Cluster-based-Graph-level strategy the calculation of the importances of the nodes was limited to the largest community's members from the initial community detection. Here we examined how the information from the largest community can effect the disintegration dynamic of the whole network. In the Cluster-level strategy the whole process was limited to the largest community of the network based on the initial community detection. In this case we were interested only in the behavior of the cluster.

3. Experiments

3.1. Amazon product co-purchasing network

Figure 1 shows the number of communities in the network as a function of the elimination step. The initial community number is 14905 based on Walktrap community detection method. In case of random elimination strategy the number of communities do not increase in elimination time, while in case of the degree- and Pagerank-based elimination strategy a linearly increasing community number can be observed. This linear effect is a bit stronger in case of Pagerank-based elimination, where the slope of the fitted linear is 17.378, while 16.581 in case of degree-based elimination.

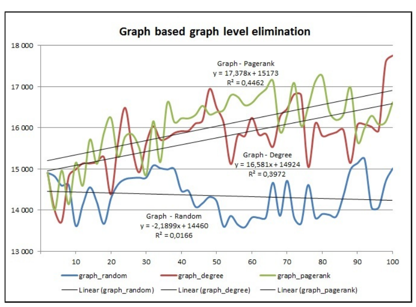

Figure 1. Graph-level elimination in Amazon product co-purchasing network

Figure 2 shows the effect of cluster based elimination strategies. Neither Pagerank-based nor degree-based elimination has effect on increasing the community number of the graph during the 100 iterations, even the 10 most important nodes were removed in each iteration step. The fitted linear function has -0.9844 slope in case of Pagerank-based elimination and -7.5856 in case of degree-based elimination. During the examination the community based information does not effect the graph structure, which implies a shell-like behavior of the community where the information from the community's inside cannot reach the other parts of the graph.

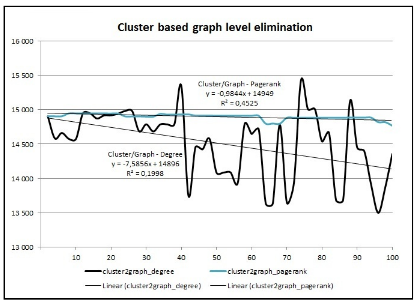

Figure 2. Cluster-based Graph-level elimination in Amazon product co-purchasing network

In case of cluster-based cluster-level elimination strategy the dynamic of disintegration is similar to the graph-based graph-level strategy's dynamic based on Figure 3. Both degree- and Pagerank-based elimination has a linear effect on disintegration of the cluster. The fitted linear has a slope 23.441 in case of Pagerank-based elimination, while 20.877 in case of degree-based elimination.

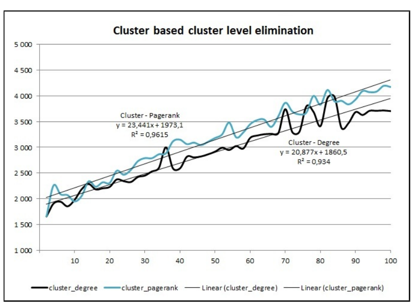

Figure 3. Cluster-level elimination in Amazon product co-purchasing network

Table 4 shows the slope of the fitted linear on community number as a function of elimination time.

Method	Ranking	Slope
Graph level	Pagerank	$17,3775$
Graph level	Degree	$16,5812$
Graph level	Random	$-2,1899$
Cluster-based Graph-level elimination	Pagerank	-0.9844
Cluster-based Graph-level elimination	Degree	-7.5856
Cluster level	Pagerank	$23,4414$
Cluster level	Degree	$20,8766$

Table 4. Slope of community number - elimination time function in case of different strategies: Amazon product co-purchasing network

3.2. Condense Matter collaboration network

The initial community number is 2514 based on Walktrap community detection method. Figure 4 shows that in case of random elimination strategy the number of communities does not increase in elimination time, while in case of the degree- and Pagerank-based elimination strategy a linearly increasing community number can be observed. This linear effect is stronger in case of Pagerank-based elimination, where the slope of the fitted linear is 8.3908, while 2.4908 in case of degree-based elimination.

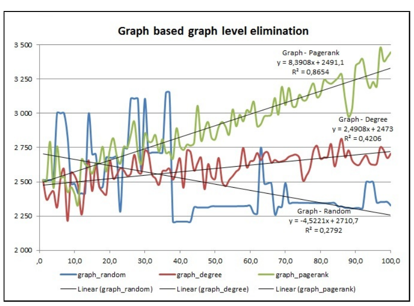

Figure 4. Graph-level elimination in Condense Matter collaboration network

Figure 5 shows the effect of cluster based elimination strategies. Neither

Pagerank-based nor degree-based elimination has effect on increasing the community number of the graph. The fitted linear function has 0.5944 slope in case of Pagerank-based elimination and 2.2707 in case of degree-based elimination. It can be also stated, that the community based information does not effect the graph structure, which also implies shell-like behavior of the community.

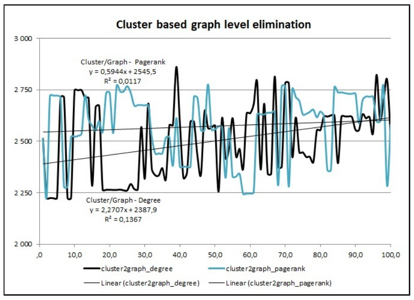

Figure 5. Cluster-based Graph-level elimination in Condense Matter collaboration network

The cluster-based cluster-level elimination strategy's dynamic of disintegration can be seen on Figure 6. Both degree- and Pagerank-based elimination has a linear effect on disintegration of the cluster. The fitted linear has a slope 0.2908 in case of Pagerank-based elimination, while 0.0803 in case of degree-based elimination.

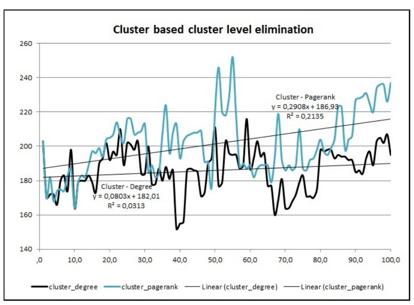

Figure 6. Cluster-level elimination in Condense Matter collaboration network

Table 5 shows the slope of the fitted linear on community number as a function of elimination time.

Method	Ranking	Slope
Graph level	Pagerank	$8,3908$
Graph level	Degree	$2,4908$
Graph level	Random	$-4,5221$
Cluster-based Graph-level elimination	Pagerank	$0,5944$
Cluster-based Graph-level elimination	Degree	$2,2707$
Cluster level	Pagerank	$0,2908$
Cluster level	Degree	$0,0803$

Table 5. Slope of community number - elimination time function in case of different strategies: CONDENSE MATTER COLLABORATION NETWORK

3.3. High-energy physics citation network

The initial community number is 710 based on Walktrap community detection method. Figure 7 shows that Pagerank- and degree-based elimination strategy has got a much larger linear effect on community number in elimination time than in case of random elimination strategy. The slope of the fitted linear is 3.5994 in case of Pagerank-based, while 2.2963 in case of degree-based elimination.

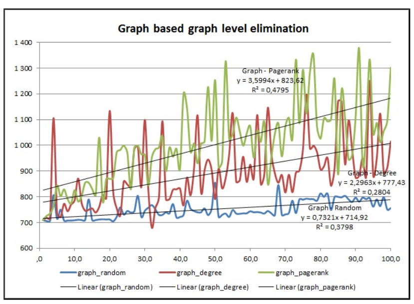

Figure 7. Graph-level elimination in HIGH-ENERGY PHYSICS CITATION NETWORK

Figure 8 shows the effect of cluster based elimination strategies. Neither Pagerank-based nor degree-based elimination has strong effect on increasing the community number of the graph. The fitted linear function has a 0.9177 slope in case of Pagerank-based elimination and 0.6635 in case of degree-based elimination. The shell-like behavior of the largest community can be obtained in this example as well.

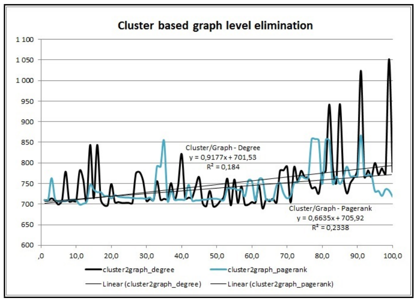

Figure 8. Cluster-based Graph-level elimination in HIGH-ENERGY PHYSICS CITATION NETWORK

In case of cluster-based cluster-level elimination strategy the dynamic of disintegration is similar to the graph-based graph-level strategy's dynamic based on Figure 9. Both degree- and Pagerank-based elimination have a linear effect on disintegration of the cluster. The fitted linear has a slope 1.9531 in case of Pagerank-based elimination, while 1.8104 in case of degree-based elimination.

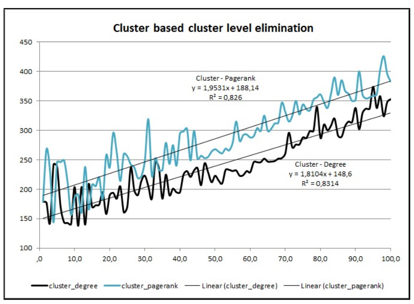

Figure 9. Cluster-level elimination in HIGH-ENERGY PHYSICS CITATION NETWORK

Table 6 shows the slope of the fitted linear on community number as a function of elimination time.

4. Summary and future plans

In this paper we have used 3 different social networks for analyzing the dynamical effect of node removal. We have shown that the disintegration of

Method	Ranking	Slope
Graph level	Pagerank	$3,5995$
Graph level	Degree	$2,2963$
Graph level	Random	$0,7321$
Cluster-based Graph-level elimination	Pagerank	$0,6635$
Cluster-based Graph-level elimination	Degree	$0,9177$
Cluster level	Pagerank	$0,6635$
Cluster level	Degree	$0,9177$

Table 6. Slope of community number - elimination time function in case of different strategies: HIGH-ENERGY PHYSICS CITATION NETWORK

the network depends on the vertex-ranking method used. For this purpose we have used random selection, degree- and Pagerank-based selection methods. To examine the role of communities in the network the walktrap community detection method was used. We have found that communities have a shell-like behavior, and it is relevant to take into account the community membership information due to its effect on disintegration dynamic during node removals. From our results it can be seen, that the effect of the node removal is different in case of different networks. One goal is to find those characteristics of the networks which are related to the choice of the proper node-ranking from disintegration point of view. Our long-term plan is to examine the role of communities in node removal strategies which can lead to the fastest disintegration of the network.

References

[1] **Albertand, R., and A.L. Barabasi,** Statistical mechanics of complex networks, *Reviews of Modern Physics,* **74** (2002), 47-97.

[2] **Boldi, P., M. Rosa, and S. Vigna,** Robustness of social and web graphs to node removal, *Social Network Analysis and Mining,* **3** (4) (2013), 829-842.

[3] **Mislove, A., M. Marcon, K.P. Gummadi, P. Druschel, P. and E. Bhattacharjee,** Measurement and analysis of online social networks, *Proc. 5th ACM/USENIX Internet Measurement Conference IMC07,* 2007.

[4] **Yan, E. and Y. Ding,** Applying centrality measures to impact analysis: A coauthorship network analysis, *J. Am. Soc. Inf. Sci.,* **60** (2009), 2107-2118. `doi: 10.1002/asi.21128`

[5] **Yang, J. and J. Leskovec,** *Defining and Evaluating Network Communities based on Ground-truth,* ICDM, 2012.

[6] **Leskovec, J., J. Kleinberg, and C. Faloutsos,** Graph evolution: Densification and shrinking diameters, *ACM Transactions on Knowledge Discovery from Data,* **1** (1) (2007).

[7] **Leskovec, J., J. Kleinberg, and C. Faloutsos,** Graphs over time: Densification laws, shrinking diameters and possible explanations, *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining,* 2005.

[8] **Gehrke, J., P. Ginsparg, and J.M. Kleinberg,** Overview of the 2003 KDD Cup, *SIGKDD Explorations,* **5** (2) (2003), 149-151.

[9] **Csardi, G. and T. Nepusz,** The igraph software package for complex network research, *InterJournal, Complex Systems 1695,* (2006).

[10] **Pons, P., M. Latapy,** *Computing communities in large networks using random walks,* http://arxiv.org/abs/physics/0512106/, 2005.

[11] **Brin, S. and L. Page,** *The anatomy of a large-scale hypertextual Web search engine, Proc. 7th World-Wide Web Conference, Brisbane, Australia, 1998.*

Imre Szücs
Eötvös Loránd University
Budapest, Hungary
icsusz@gmail.com

Attila Kiss
Eötvös Loránd University
Budapest, Hungary
kiss@inf.elte.hu