

Structural Micro Forces in Flickr Social Network

Haris Memic*
Faculty of Engineering and Natural Sciences, International University of Sarajevo
Sarajevo71000, Bosnia and Herzegovina
E-mail: memich@gmail.com

Alen Savatic
Faculty of Engineering and Natural Sciences, International University of Sarajevo
Sarajevo71000, Bosnia and Herzegovina
E-mail: alen@savatic.net

Abstract
Previous studies on network structure of large online social networks focused almost exclusively on exploring the global network descriptive statistics. Few of the studies that have researched how these networks evolve from the micro forces at the local level have fallen short of modeling reciprocity and different ways triangle closure. By focusing on the denser areas of the Flickr network (user groups) and with the help of recently extended biased net modeling framework, we specified and fitted models and estimated parameters for all possible purely structural dyadic and triadic network effects: reciprocity, transitivity, structural similarity, closure and cyclicality. Our results showed that the reciprocity is by far the most strongest force acting in the network, followed by transitivity, closure and structural similarity. Cyclicality has been, surprisingly, proven not to exist at all. Furthermore, we have found that the size of the groups corresponds negatively with the magnitude of each of the micro forces.
Keywords: Flickr, online networks, network structure, online social networks

1. Introduction

The structural network properties of online social networks have been examined extensively during the last couple of years. However, the focuses of most research studies have been global, general network structural properties such as diameter, clustering coefficients and degree distributions. The focal point has almost exclusively been on the macro level analysis of the structure of online social networks. The micro network effects, or the forces acting on the micro level of these networks, on another hand, have either not been studied or have not been studied well [1]. This is most likely due to do the fact that with the available computing resources it is impossible to fit statistically reliable computational models which would in realistic time be able to provide us with parameter estimates for large networks such are usually online social networks [2, 3]. The small number of studies that examined micro network forces is of a very limited value, as is elaborated in the next section (literature review) of the paper. The same applies to Flickr social network, whose data we used in this study of micro level structural forces. In order to be able to proceed where past studies did not, we adopted the strategy proposed in [4], namely to analyze appropriate subgroups of large online social networks, so as to be able to computationally fit models to observed network data and to come to valuable results.

More specifically, our goal in this research was to study the micro forces at the most atomic levels: dyadic and triadic levels. The long history of the sociological side of social network analysis based on numerous empirical findings pertaining to offline human social networks [e.g. 5, 6], has shown that these forces are usually almost always the most significant forces acting in human social networks, together with the force of homophily. Moreover, the above mentioned studies have strongly indicated that these are the major forces whose agglomeration over the time causes the creation of the global network structure. This is also the hypothesis taking in this study. Consequently, we decided to examine the dyadic force of reciprocity, and all of the triadic structural forces: transitivity, closure, structural similarity and cyclicality.

The reciprocity force corresponds to the tendency of reciprocating an incoming link. Four triadic forces are illustrated in Fig. 1 and correspond to four different ways that a triangle can be closed.

Our modeling strategy was inspired by the biased net modeling framework as proposed in [7], but with the recent extension related to incorporation of cyclicality as proposed in [8, 9].

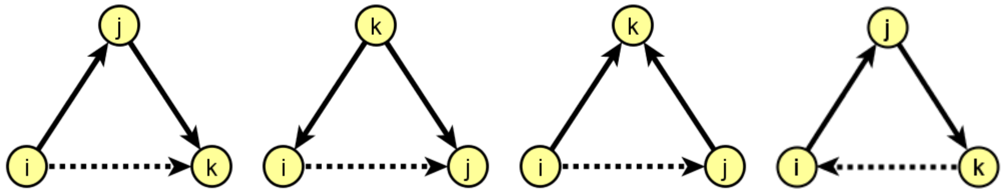

Figure 1. Four triadic forces illustrated with doted links. From left to right: closure, transitivity, cyclicality, structural similarity.

The rest of the paper is organized in the following way. In section 2, literature review of relevant studies is provided. In section 3 we explain the statistical modeling methodology used in the study. Section 4 presents the network data used in this research together with the strategy we decided to follow in our data analyses. Section 5 presents the results of our research and section 6 discusses and summarizes the findings.

2. Literature Review

Past research efforts on network structure of online social networks examined a variety of global network properties of these networks by using a number of descriptive methods for network analysis. The majority of these studies examined degree distribution, degree correlation and clustering coefficients of online social networks. Some of the more prominent such studies are [10, 11, 12, 13, 14]. Other studies used novel approaches to study "untraditional" global network measures such as density distribution [15], network segmentation [16], or similar other global network characteristics.

Few studies researched micro-level structural forces in large online social networks, focusing primarily on reciprocity [17, 18]. These are, however, of a very limited statistical value, as they studied individual micro forces in isolation of other local forces. For that reason, as an example, many reciprocated dyads could by the byproduct of the very dense network areas, formed by a strong clustering mechanism.

The breakthrough work in the analysis of micro forces was [19] which proposed a model for evolution of online social networks. Model incorporated a triangle closure mechanism by studying different ways of triangle closure. Although valuable, this research was limited in its design, as it didn't take into account the link direction and thus was unable to differentiate between different ways of triangulation (transitivity, closure, similarity, cyclicality), nor was it able to model the reciprocity.

3. Modeling Framework

The recently proposed extension of the biased net modeling framework [8, 9], incorporating the bias of cyclicality, allowed us to examine all of the simple triadic forces together with the reciprocity.

Biased net statistical framework associates micro structural processes (forces) with hypothetical bias stochastic events such that if a certain bias event happens, a particular kind of a network link forms with certainty (with Pr=1). Relevant bias events for the specific link/dyad depend on the rest of the network.

It is assumed that if all of the possible bias events fail, the link still may still form with a random chance of connection, which basically sums up all of the non-modeled network forces. It is further assumed that all bias events and the random chance of connection are probabilistically independent events. Thus, the probability that a link fails to happen is the product of probabilities of all relevant bias events for that link failing to happen.

As an example, the potential link from node 4 towards node 2, in Fig. 2., is subject to multiple bias events: a) reciprocity event, as there is already link 2->4, b) cyclicality event, as there are links 2->1 and 1->4, and c) closure event, as there are links 1->2 and 1->4.

The model is formulated so that its parameters represent probabilities of individual micro processes (bias events) in social networks.

The probability that all events fail to occur is the product of the failure probabilities of each event. Consequently, the probability of the link forming then corresponds to (1).

1 – (probability all events fail to occur). (1)

Models specify how the outcomes in a dyad depend on the bias events and the random chance of connection, conditional on the rest of the network, as illustrated with following equations.

$$
\begin{aligned}
&\Pr\left(D_{ij00} = 1 \middle| G^{-ij}\right) \\
&\quad = (1-c)^{m_{ij}}(1-t)^{t_{ij}}(1-s)^{s_{ij}}(1-b)^{t_{ji}}(1-d) \\
&\qquad * (1-c)^{m_{ij}}(1-t)^{t_{ji}}(1-s)^{s_{ij}}(1-b)^{t_{ij}}(1-d) \\
&\Pr\left(D_{ij01} = 1 \middle| G^{-ij}\right) \\
&\quad = [1-(1-c)^{m_{ij}}(1-t)^{t_{ji}}(1-s)^{s_{ij}}(1-b)^{t_{ij}}(1-d)] \\
&\qquad * (1-r)(1-c)^{m_{ij}}(1-t)^{t_{ij}}(1-s)^{s_{ij}}(1-b)^{t_{ji}}(1-d) \\
&\Pr\left(D_{ij10} = 1 \middle| G^{-ij}\right) \\
&\quad = [1-(1-c)^{m_{ij}}(1-t)^{t_{ij}}(1-s)^{s_{ij}}(1-b)^{t_{ji}}(1-d) \\
&\qquad * (1-r)(1-c)^{m_{ij}}(1-t)^{t_{ji}}(1-s)^{s_{ij}}(1-b)^{t_{ij}}(1-d) \\
&\Pr\left(D_{ij11} = 1 \middle| G^{-ij}\right) \\
&\quad = 1 - \Pr\left(D_{ij00} = 1 \middle| G^{-ij}\right) - \Pr\left(D_{ij01} = 1 \middle| G^{-ij}\right) - \Pr\left(D_{ij10} = 1 \middle| G^{-ij}\right)
\end{aligned}
\tag{2}
$$

In the above equations: m_{ij} refers to the number of common nodes that have link to both *i* and *j* (risks for closure); t_{ij} refers to the number of nodes toward whom *i* has link and from whom there is a link to *j* (risks for transitivity); s_{ij} refers to the number of nodes toward which both *i* and *j* have links (risks for structural similarity); t_{ji} refers to the number of nodes toward whom *j* has link and from whom there is a link to *i* (risks for cyclicality).

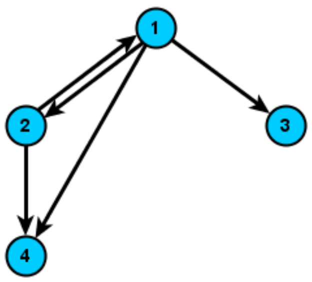

Figure 2. A sample network illustrating multiple possible bias events for link from node 4 to node 2.

D_{ijuv} denotes a random variable which corresponds to the state of *ij* dyad. D_{ij01} for example, denotes a dyadic state in which there is no link from *i* to *j*, but in which there is a link from *j* to *i*.

Parameter notation is as follows: density (d), reciprocity (r), transitivity (t), closure (c), cyclicality (b) and similarity (s).

More detailed description of the modeling approach and pseudo-likelihood method for parameter estimation can be found in our other works [20, 7, 9].

4. Data and Data Analysis Strategy

Flickr is a popular image and video hosting online community and a social network, known almost exclusively as an image sharing platform.

The Flickr social network data was provided to us from Alan Mislove and was collected in January 2007. It contains over 1.8 million users and 22 million network links and encompassed the large proportion of users belonging to the main weakly connected component of the Flickr network. More detailed description of the data and collection methods can be found in [21]. One of the authors of this paper wrote scripts to extract the network data from the provided data files, and to convert the extracted data to the matrix format as suitable for biased net modeling.

Flickr users have option to join various interest-based groups. In [21] it was found that group sizes follow a power-law distribution and that a large majority of the groups have only few users. Flickr users are also able to create direct links with other users. It was found that user groups are more heavily connected areas of Flickr social network, when compared to the other parts of the network, and that smaller groups tend to be more clustered than the larger ones [21].

Because the detection of statistical significance tends to be easier for larger groups, and while modeling micro level biases (forces) in networks of more than few hundred nodes tends to be an exceptionally slow process even on supercomputers, our strategy was to focus our analyses on smaller groups.

Very small groups such as those bellow 50 users tend to be both less representative of the overall Flickr network structure and more fragile and inconsistent than the larger groups, which is the reason why we omitted them from our study.

As one of the goals of our research was to study whether the presence and/or strength of individual micro level forces depends on the size of the groups, we have divided groups in nine subgroupings based on the number of group members: subgrouping1: groups who had between 50 and 59 members, subgrouping2: 60-69 members, subgrouping3: 70-79, subgrouping4: 80-89, subgrouping5: 90-99, subgrouping6: 100-109, subgrouping7: 110-119, subgrouping8: 120-129, and subgrouping9: 130-139 members. From the provided data we extracted a total of 2000 of groups with sizes between 50 and 139 members/users.

From obtained groups, we selected 200 of groups for each of the subgroupings, in a manner so that the sizes of these groups depart as less as possible from artificial means of the nine subgroupings, which we selected to be: 55, 65, 75, 85, 95, 105, 115, 125 and 135, and then applied a mean crowding so as to ensure that means of the subgroupings correspond as closely as possible to our aimed means.

Our achieved means are as shown in Table 1.

Table 1. Achieved means for subgroupings

Subgr.1 Mean: 55.1	Subgr.2 Mean: 65.0	Subgr.3 Mean: 75.0
Subgr.4 Mean: 85.0	Subgr.5 Mean: 95.0	Subgr.6 Mean: 105.0
Subgr.7 Mean: 115.0	Subgr.8 Mean: 124.4	Subgr.9 Mean: 135.0

5. Results

Results of the fitted models which incorporated density, reciprocity and all of the triadic forces in the same models, thus controlling for each other, are presented bellow.

A. Density

Parameter for the density was estimated to be significant in almost all of the networks. The number of significant, non-significant, and non-estimable (NaN) density parameters for each of the 9 subgroupings (each consisting of 200 Flickr groups), is shown in the Table 2.

Table 2. Significance and magnitude of density

Subgroup	Avg Density	Significant	Insignificant	NaN	Total groups
55.1	0.0079	191	6	3	200
65.0	0.0069	193	5	2	200
75.0	0.0063	195	3	2	200
85.0	0.0055	192	4	4	200
95.0	0.0051	193	6	1	200
105.0	0.0048	198	0	2	200
115.0	0.0043	191	2	7	200
124.4	0.0042	191	4	5	200
135.0	0.0039	190	9	1	200

Average values of estimated parameters for the density, for each of the subgroupings, are also shown in the above table, and are further displayed in Fig. 3.

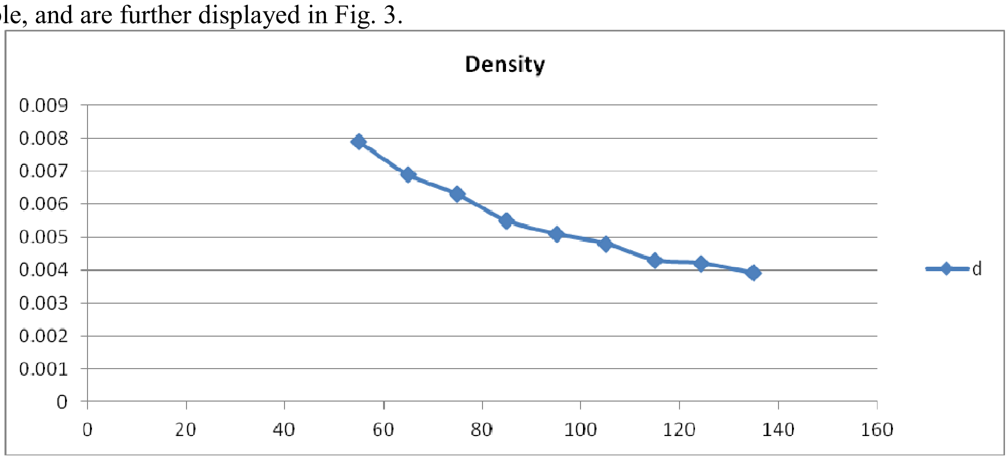

Figure 3. Average values for density, for each of the subgroupings.

B. Reciprocity

Even though the reciprocity is present in most of the human offline social networks, the surprising result of our modeling is that in absolutely all of the 1800 examined Flickr groups reciprocity was proved to be statistically significant at the 0.05 level of statistical significance. These results, together with averaged values of reciprocity parameters, corresponding to the probability of reciprocity bias events happening, are showed in the Table 3.

Table 3. Significance and magnitude of reciprocity

Subgroup	Avg Recipr.	Significant	Insignificant	NaN	Total groups
55.1	0.6024	200	0	0	200
65.0	0.5881	200	0	0	200
75.0	0.5724	200	0	0	200
85.0	0.5655	200	0	0	200
95.0	0.5686	200	0	0	200
105.0	0.5592	200	0	0	200
115.0	0.5455	200	0	0	200
124.4	0.5548	200	0	0	200
135.0	0.5371	200	0	0	200

Averaged values of reciprocity for each of the subgroupings are also displayed in the Fig. 4.

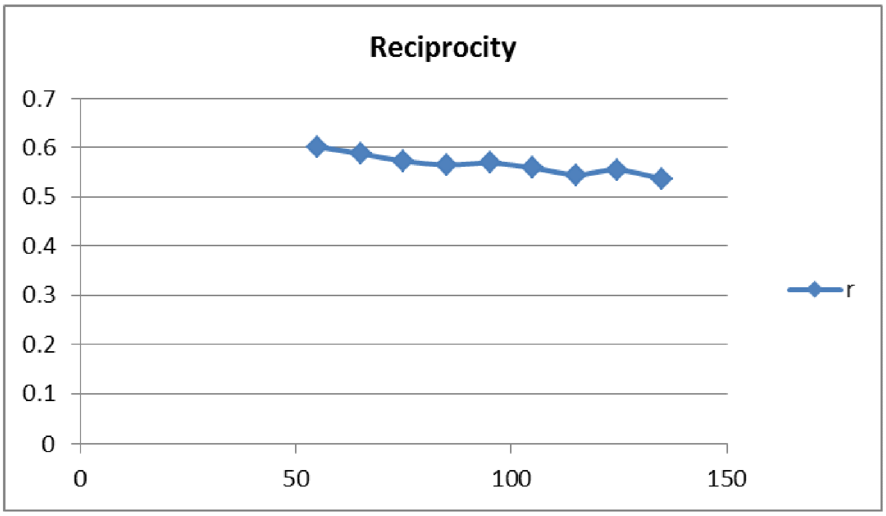

Figure 4. Average values for reciprocity, for each of the subgroupings.

C. Closure

We have found statistically significant presence of closure in 16% of the groups in which we were able to obtain estimable parameters with this percentage raising to 29% for the subgrouping9. Results are presented in Table 4. The average values of closure parameters (probability biases) for different subgroupings are showed in the same table, and are additionally plotted in Fig. 5.

Table 4. Significance and magnitude of closure

Subgroup	Avg Closure	Significant	Insignificant	NaN	Total groups
55.1	0.0176	6	184	10	200
65.0	0.0141	17	174	9	200
75.0	0.0107	24	165	11	200
85.0	0.0098	32	155	13	200
95.0	0.0088	28	164	8	200
105.0	0.0092	28	161	11	200
115.0	0.0080	34	146	20	200
124.4	0.0075	45	140	15	200
135.0	0.0066	55	133	12	200

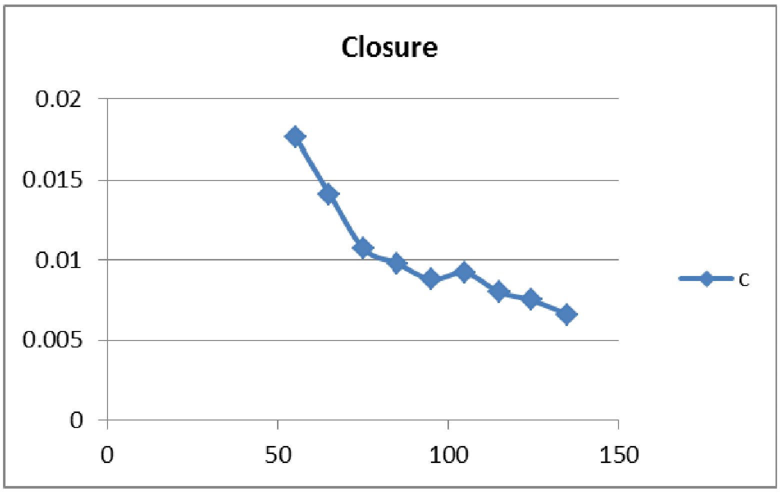

Figure 5. Average values for reciprocity, for each of the subgroupings.

D. Transitivity

We have found statistically significant presence of transitivity in the significant majority of groups, especially in the larger ones. These results, together with mean values of estimated parameters for transitivity, corresponding to values of transitivity force, for each of the subgroupings are shown in Table 5.

Table 5. Significance and magnitude of transitivity

Subgroup	Avg Transit.	Significant	Insignificant	NaN	Total groups
55.1	0.0513	80	111	9	200
65.0	0.0516	104	86	10	200
75.0	0.0496	132	59	9	200
85.0	0.0484	136	55	9	200
95.0	0.0441	152	41	7	200
105.0	0.0473	151	42	7	200
115.0	0.0445	154	35	11	200
124.4	0.0395	167	23	10	200
135.0	0.0397	181	15	4	200

Fig. 6 displays the mean values of transitivity parameters for each of the subgroupings.

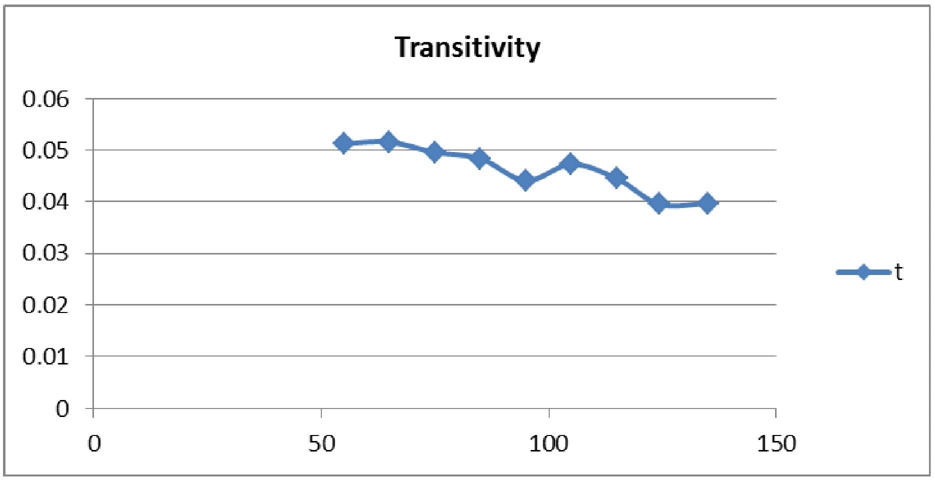

Figure 6. Average values for transitivity, for each of the subgroupings.

E. Structural Similarity

With regards to the structural similarity, we found it on average to be significant in 12% of the groups (in which parameter estimation was possible), but this number increased to 25% for the subgrouping9. Table 6 summarizes these results and also shows the average values of structural similarity parameters for each of the subgroupings, corresponding to probabilities of structural similarity force/bias. These are also plotted in Fig. 7.

Table 6. Significance and magnitude of structural similarity

Subgroup	Avg Str. Similarity	Significant	Insignificant	NaN	Total groups
55.1	0.0513	80	111	9	200
65.0	0.0516	104	86	10	200
75.0	0.0496	132	59	9	200
85.0	0.0484	136	55	9	200
95.0	0.0441	152	41	7	200
105.0	0.0473	151	42	7	200
115.0	0.0445	154	35	11	200
124.4	0.0395	167	23	10	200
135.0	0.0397	181	15	4	200

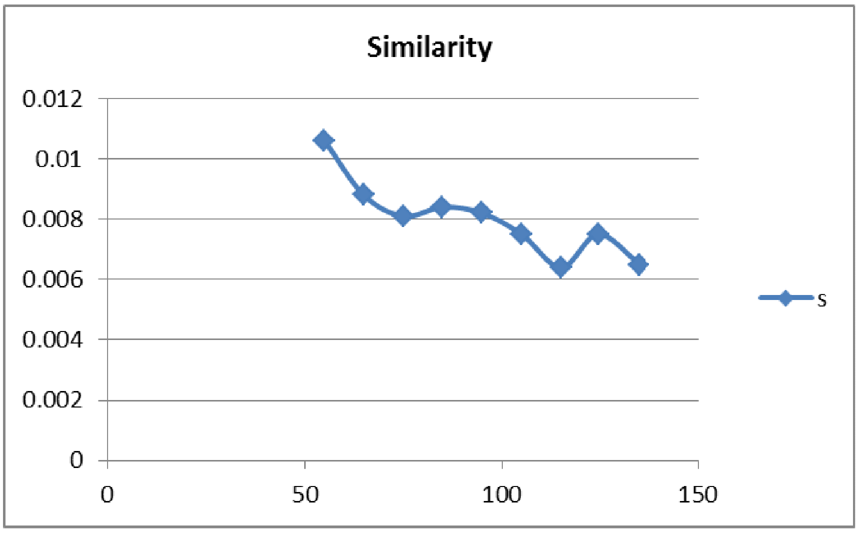

Figure 7. Average values for structural similarity, for each of the subgroupings.

F. Cyclicality

We found that cyclicality force is virtually totally absent from Flickr groups. In all but one single group it was found to be insignificant, as shown in the Table 7.

Subgroup	Avg Cyclic.	Significant	Insignificant	NaN	Total groups
55.1	0.0057	0	184	16	200
65.0	0.0032	0	173	27	200
75.0	0.0028	0	164	36	200
85.0	0.0017	1	152	47	200
95.0	0.0014	0	160	40	200
105.0	0.0012	0	155	45	200
115.0	0.0010	0	150	50	200
124.4	0.0008	0	165	35	200
135.0	0.0005	0	146	54	200

The average values of parameters for the force of cyclicality are also presented in the above table and are plotted in Fig. 8.

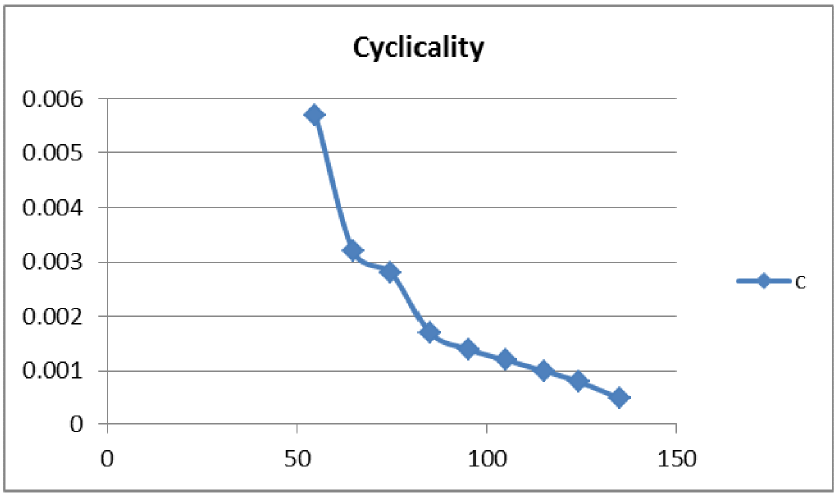

Figure 8. Average values for cyclicality, for each of the subgroupings.

6. Discussion and Concluding Remarks

This work studied how the network structure of a large online social network Flickr develops by agglomeration of forces which act on local network levels. In order to proceed beyond analysis of global network properties, which we hypothesized are the byproducts of what is happening on the local network levels, we focused on analysis of the denser parts of the overall Flickr network, on its user groups whose sizes are dramatically smaller than the overall network. This allowed us to apply the extended version of the biased net modeling framework to estimate parameters for each of the hypothesized purely structural forces on the dyadic and triadic network levels. Results of our statistical modeling of 1800 Flickr groups showed that the reciprocity is by far the most important force acting in the network, regardless of the size of groups. Transitivity was the second most important local force acting in the network, both in terms of statistical significance and in terms of the strength. The third most significant structural micro bias acting in the Flickr is closure, followed by structural similarity, again both in terms of statistical significance and the magnitude. Both of these however are much less significant and strong than the transitivity. We have found no traces of cyclicality, which in itself is a fascinating discovery.

Additionally, we found group sizes to be positively correlated with the significance of the forces, excluding reciprocity and cyclicality which are always significant / respectively insignificant. It remains to be investigated whether this correlation has to do with the statistically known fact that the detection of statistical significance tends to be easier when the number of observations is larger, or is perhaps related to some other phenomena.

Surprisingly, we have also found that the size of the groups corresponds negatively with the magnitude of each of the micro forces. This opens room for further interesting investigations.

Acknowledgment

The authors wish to thank Alan Mislove and his team for providing us with Flickr data.

References

[1] H. Memic, “Structural Micro Forces in Online Social Networking Websites: Impact on Friendship Structure,” in Proc. 3rd IEEE International Conference on Digital Ecosystems and Technologies (DEST 2009), IEEE Xplore, 2009, pp. 390-395.

[2] H. Memic, “Structure in Online Social Networks: Bridging the Micro-Macro Gap,” in Proc. 31st International Conference on Information Technology Interfaces (ITI 2009), IEEE Xplore, 2009, pp. 267-272.

[3] S. M. Goodreau, “Advances in Exponential Random Graph (p*) Models Applied to a Large Social Network,” Social Networks, vol. 29(2), pp. 231-248, June 2007.

[4] H. Memic, “Triadic Patterns of Friendships in YouTube Groups,” in Proc. 12th IEEE International Symposium on Computational Intelligence and Informatics (CINTI 2011), IEEE Xplore, 2011, pp. 501-506.

[5] J. G. Scott and P. J. Carrington, The SAGE Handbook of Social Network Analysis, Thousand Oaks, CA: Sage Publications, 2011.

[6] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences, Cambridge, U.K.: Cambridge University Press, 1994.

[7] J. Skvoretz, “A Reformulated Family of Biased Net Models,” Unpublished manuscript.

[8] H. Memic and J. Skvoretz, “Models for Micro-forces: Application to an Online Friendship Network,” presented at 33rd International Sunbelt Social Network Conference, Hamburg, Germany, May 21-26, 2013.

[9] H. Memic, "Micro Structural Processes in Online Social Networks," Unpublished, 2013.
[10] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in Proc 5th ACM/USENIX Internet Measurement Conference, 2007, pp. 29-42.
[11] P. Holme, C. R. Edling and F. Liljeros, "Structure and Time-Evolution of an Internet Dating Community," Social Networks, vol. 26, 2004, pp. 155-174.
[12] K. Yuta, K. N. Ono and Y. Fujiwara, "A gap in the community-size distribution of a large scale social networking site," arXiv.org: physics/0701168, 2007.
[13] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan and A. Tomkins, "Geographic Routing in Social Networks," in Proc. National Academy of Sciences (PNAS), vol. 102 no. 33, 2005, pp. 11623–11628.
[14] R. Kumar, J. Novak and A. Tomkins, "Structure and Evolution of Online Social Networks," in Proc. 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 611-617.
[15] D. Zinoviev, "Topology and Geometry of Online Social Networks," in Proc. 12th World Multi-Conference on Systemics, Cybernetics and Informatics, 2008, pp. 138-143.
[16] Y. Y. Ahn, S. Han, H. Kwak, S. Moon and H. Jeong, "Analysis of Topological Characteristics of Huge Online Social Networking Services," in Proc. 16th international conference on World Wide Web, 2007, pp. 835-844.
[17] H. Chun, H. Kwak, Y. H. Eom, Y. Y. Ahn, S. Moon and H. Jeong, "Comparison of Online Social Relations in Volume vs Interaction: a Case Study of Cyworld," in Proc. 8th ACM SIGCOMM conference on Internet measurement, 2008, pp. 57-70.
[18] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel and B. Bhattacharjee, "Growth of the Flickr social network," in Proc. 1st Workshop on Online Social Networks, 2008, pp. 25-30.
[19] J. Leskovec, L. Backstrom, R. Kumar and A. Tomkins, "Microscopic Evolution of Social Networks," in Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 462-470.
[20] H. Memic, "Structural Micro Forces in Online Social Networking Websites: Impact on Friendship Structure", in Proc. 3rd IEEE International Conference on Digital Ecosystems and Technologies (DEST 2009), IEEE Xplore, 2009, pp. 390-395.
[21] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," Proc. 5th ACM/USENIX Internet Measurement Conference (IMC'07), 2007, pp. 29-42.