Individual Neighbourhood Exploration in Complex Multi-layered Social Network

Przemysław Kazienko[1], Piotr Brodka[1], Katarzyna Musial[2,3],
[1] Institute of Informatics, Wrocław University of Technology
Wyb.Wyspiańskiego 27, 50-370 Wrocław, Poland
[2] School of Design, Engineering & Computing, Bournemouth University, Poole, Dorset, BH12 5BB, United Kingdom
[3]Telnet Sp. z o.o., Obrońców Poczty Gdańskiej 13a, 52-204 Wrocław, Poland
kazienko@pwr.wroc.pl, piotr.brodka@pwr.wroc.pl, kmusial@bournemouth.ac.uk

Abstract

Social networks can be extracted from different data about communication or common activities in organizations, companies or various Internet-based services. Different types of data processed may result in creation of separate layers in the complex multi-layered social network. Analysis of neighbourhoods of network members and their utilization to social group discovery appears to be an interesting and important research domain. Since there is no measure to evaluate structure of the neighbourhoods in the multi-layered social network, a new measure called cross layered multi-layered clustering coefficient (CLMCC) is proposed in the paper. It enables to analyse the density of mutual connections of neighbours that occur in at least a given number of layers in a social network. Additionally, experimental studies on real-world data are presented.

1. Introduction

The area of complex networks [1] has attracted more and more scientists from different research fields for last few decades. All complex network systems have some common features such as: (i) skewed distribution of connections, (ii) small degree of separation between vertices, (iii) non-trivial temporal evolution, (iv) relatively high clustering rate (v) and presence of motifs, hierarchies and communities [3], [7]. The type of complex system that is analysed by us is a social network. It is a structure that consists of nodes (people) and edges that denote relationships between people [4], [8]. This article focuses on the analysis of neighbourhoods within the complex multi-layered social networks (CMSN) extracted from IT systems, in which users interact or cooperate with each other by means of various dedicated services. The main profile of CMSN is that it consists of many layers, corresponding to different kinds of relationship [5]. The analysis of neighbourhoods in social networks is not a new scientific domain [6], [9]. However, none of the research addresses this issue for complex multi-layered social networks CMSN.

2. Complex Multi-layered Social Network

The structure that will be analysed in this research is *a social network*, i.e. a set of interconnected nodes [8]. Social networks are one of the subgroup of complex networked systems [3], in which nodes are social entities and the connections between them usually reflect social interactions between these entities. In this paper, the social networks extracted from different systems based on communication technologies will be investigated. The units (nodes) are digital representations of people who use email services, telecommunication systems, multimedia sharing systems, access blogosphere etc. Based on interactions between users and their activities within the system the relationships are extracted. Due to diversity of communication channels the analyzed networks are *multi-layered,* i.e. they consist of more than one type of relationship. Different relations can emerge from different communication channels, i.e. based on each communication channel separate relation is created. It is called a layer of a network. The described above concept is called a complex multi-layered social network and its formal definition is as follows:

Definition 1

Complex multi-layered social network (CMSN) is defined as a tuple $<V, E, L>$
where:

- V – is a not empty set of nodes,
- E – is a set of tuples $<x,y,l>$, $x,y \epsilon V$, $l \epsilon L$, $x \neq y$ and for any two tuples $<x,y,l>$, $<x',y',l'> \epsilon E$ if $x=x'$ and $y=y'$ than $l \neq l'$. Each tuple is an edge in a complex multi-layered social network (CMSN),
- L – is a set of layers. One layer corresponds to a simple, one-layered social network.

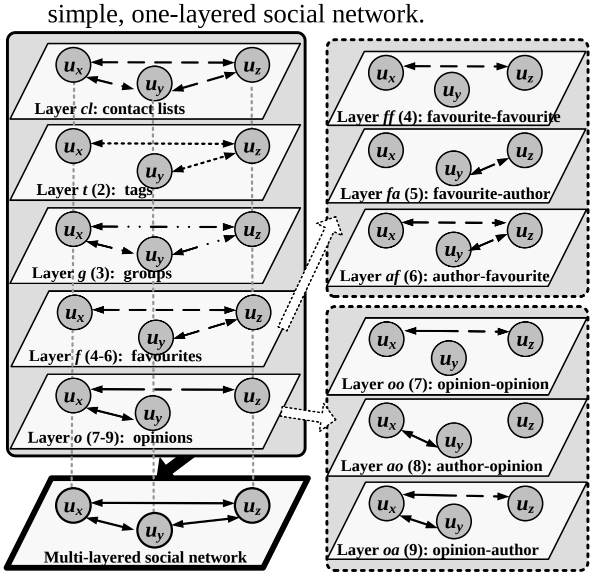

Fig. 1. The example of multi-layered complex social network extracted from Flickr system

One of the examples of CMSN is presented in Fig. 1. It shows the layers that were identified in the Flickr[1] photo sharing system. During the research presented in [5], nine types of relations were extracted in Flickr: relations based on contact lists –*contact list layer*, shared tags used by more than one user –*tags layer*, user groups –*groups layer*, photos added by users to their favourites –*favourite-favourite*, *favourite-author*, *author-favourite layers*, and opinions about pictures created by users –*opinion-opinion*, *opinion-author*, *author-opinion layers*. Relations based on contact lists (R^c) represent direct relations where users directly interact with each other. Tag-based (R^t), group-based (R^g), favourite-favourite (R^{ff}), and opinion-opinion relations (R^{oo}), favourite-author (R^{fa}), author-favourite (R^{af}), opinion-author (R^{oa}), and author-opinion (R^{ao}) are relations which are created based on activities that users performed towards some objects in the system rather than a specific person, e.g. has commented a given photo. All these relations correspond to nine separate layers in one multi-relational social network

3. Cross Layered Multi-layered Clustering Coefficient

The regular local clustering coefficient describes how close neighbours of a given node in the network structure tend to create a clique i.e. fully connected graph. Local clustering coefficient was presented by Duncan J. Watts and Steven Strogatz who used it to calculate whether a graph is a small-world network [10].

In this article, we propose a redefinition of the regular clustering coefficient, We called this new metric: a cross layer multi-layered clustering coefficient C*LMCC* which allows to calculate clustering coefficient in complex multi-layered social networks.

First we have to define the cross layered neighbourhood $CLN(x,\alpha)$ of a given node x:

$$CLN(x,\alpha) = \{ y : card(\{ < x,y,l > \in E \}) \geq \alpha \} \qquad (1)$$

Basically CLN is a set of nodes which are neighbours of node x at least in α layers in CMSN. For example $CLN(x,4)$ means that cross layered neighbourhood consist only of nodes which are neighbours of x at least on four layers.

The cross layer multi-layered clustering coefficient $CLMCC(x,\alpha)$ for a given node x, is computed in the following way:

$$CLMCC(x,\alpha) = \frac{\sum_{l \in L} \sum_{y \in CLN(x,\alpha)} \left(in(y, CLN(x,\alpha), l) + out(y, CLN(x,\alpha), l) \right)}{2 \cdot card(CLN(x,\alpha)) \cdot card(L)} \qquad (2)$$

where:

- $in(y,CLN(x,\alpha),l)$ – the weighted in-degree of node y that belongs to cross layered neighbourhood of x in the network containing one layer l,
- $out(y,CLN(x,\alpha),l)$ – the weighted out-degree of node y that belongs to the cross layered neighbourhood of x in the network containing one layer l.

The weighted in-degree $in(y, CLN(x,\alpha),l)$ for a given node y in the network containing one layer l is the sum of all weights $w(z,y,l)$ of incoming edges $<z,y,l>$ from network containing one layer l that income to a given node y from other nodes z from the neighbourhood $CLN(x,\alpha)$.

[1] http://www.flickr.com/

$$in(y, CLN(x,\alpha), l) = \sum_{z \in CLN(x,\alpha)} w(z, y, l), \quad y \in CLN(x,\alpha). \qquad (3)$$

Likewise, the weighted out-degree *out*(*y*,*CLN*(*x*,*α*),*l*) for a given node *y* is the sum of all weights *w*(*y*,*z*,*l*) of the outgoing edges <*y*,*z*,*l*> that come from *x*'s neighbours *z*.

Note that the sum of weights of outgoing edges for a given node is always 1, so *CLMCC*(*x*,*α*) is always from the range (0;1]. It equals 1, where each neighbour *y*∈*CLN*(*x*,*α*) have outgoing relationships towards all other nodes *z*∈*CLN*(*x*,*α*)

The detailed formula for multi-layered clustering coefficient (MCC) as well as the two special cases of *CLMCC* for *α=1* (MCCEN) and *α*=*card*(*L*) (MCCRN) was described in [2].

4. Experimental Study and Discussion

The main goal of the experiments was to investigate the characteristics of proposed in the paper *CLMCC* metric that serves to assess the extent to which the neighbourhoods of user are clustered within the complex multi-layered social networks.

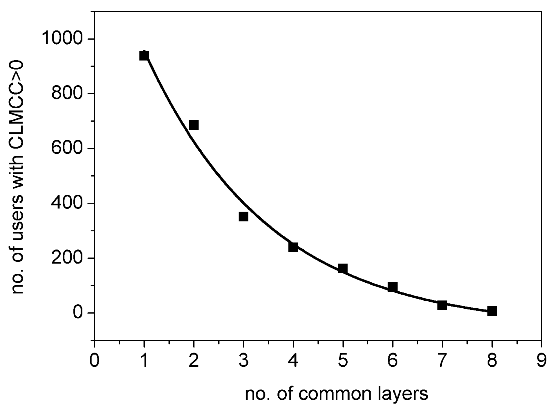

Fig. 2 Number of users with *CLMCC*>0 depending on number of layers for which *CLMCC* was calculated

The experiments were performed on 1000 users from the Flickr system where nine different layers have been identified. These layers include: tags used by more than one user R^t, user groups R^g, photos added by users to their favourites R^{fa}, R^{af}, R^{ff}, opinions about photos created by users R^{oa}, R^{ao}, R^{oo}, and the relations derived from the contact lists R^c. The detailed information about the data set can be found in [5]. However in the layer R^{ff} (*favourite-favourite*) there was no data available, so finally it was excluded and the rest eight layers were analysed.

First, we checked how many users posses *CLMCC* greater than zero in one, two, etc layers. The number of such users decreases exponentially with the number of common layers for which *CLMCC* was calculated (Fig. 2). For one layer 93.8% of users have *CLMCC* greater than 0 and for eight layers it is only 0.7% of users. The experiments have revealed that the best fitting function that describes the dependence between number of users with *CLMCC*>0 (*y*) and number of layers for which *CLMCC* was calculated (*x*) is an exponential decay function that is expressed by the formula: $y(x) = 1507.6 \cdot e^{-0.39 \cdot x} - 58.97$ (Fig. 2). The correlation rate equals 0.987.

The outcomes of the experiments have shown that the maximum value of *CLMCC* coefficient is when calculated for two layers and equals 0.576. In the same time the average value of *CLMCC* equals 0.425 for one layer and 0.281 for two layers and then substantially decreases for three and more layers (Tab. 1). When analysing standard deviation, it can be noticed that it is lower than 0.1 for almost all cases. The only exception if α=2 where it equals 0.241 (Tab. 1) so the average value in this situation is not an informative one.

No. of layers (*α*)	Max *CLMCC*	Average *CLMCC*	Std. Dev.
1	0.534	0.425	0.094
2	0.576	0.281	0.241
3	0.349	0.030	0.063
4	0.321	0.016	0.046
5	0.316	0.010	0.038
6	0.269	0.005	0.028
7	0.292	0.002	0.020
8	0.193	0.001	0.009

Tab. 1 Maximum, average, and standard deviation value of *CLMCC* for different no. of min. common layers

The tendency is that if the larger number of layers is taken for the *CLMCC* calculation then the greater number of users have the *CLMCC* equals 0 (Fig. 3). For most α values the best fitting function that describes the value of *CLMCC* for users in the Flickr system (sorted ascending by *CLMCC*) is an exponential growth (Fig. 3). Only for α=2 and α=8 it is polynomial function with degree 5 and 2 respectively (Fig. 3). It can be noticed that the *CLMCC* for eight layers equals 0 for 99.3% of users and it denotes that there are only few users (7) who have similar neighbourhoods on all layers. This denotes that users tend to maintain the relationships with different neighbourhood on different layers. This is also confirmed by the relatively high *CLMCC* value for one (average is 0.425).

The interesting difference can be noticed between the situation where *α*=1 and *α*=2. The maximum value of *CLMCC* is greater for *α*=2 than for *α*=1. This can be caused by the fact that although the neighbourhood in the case when α=2 is smaller, it is more internally connected.

5. Conclusions

The paper addresses the problem of neighbourhood analysis in the CMSN. Users can be connected with each other on different layers of CMSN. The new definition of cross-layered neighbourhood has been proposed and it consists of users whose activities result in relations on at least a given number of layers. The cross-layered multilayered clustering coefficient *CLMCC*, defined in the paper, measures the density of connections between users that belongs to the cross-layered neighbourhood (*CLN*) The experiments conducted revealed that the average value of *CLMCC* generally decreases with the increasing number of layers. Our future work will focus on further experiments and comparison of the *CLMCC* coefficient with the value of the proposed coefficient in the random multi-layered social networks.

Acknowledgments. The authors are indebted to Tomasz Kajdanowicz for the data collection delivery and Anna Musiał for valuable remarks. This work has been supported by the Polish Ministry of Science and Higher Education, the development project, 2009-11.

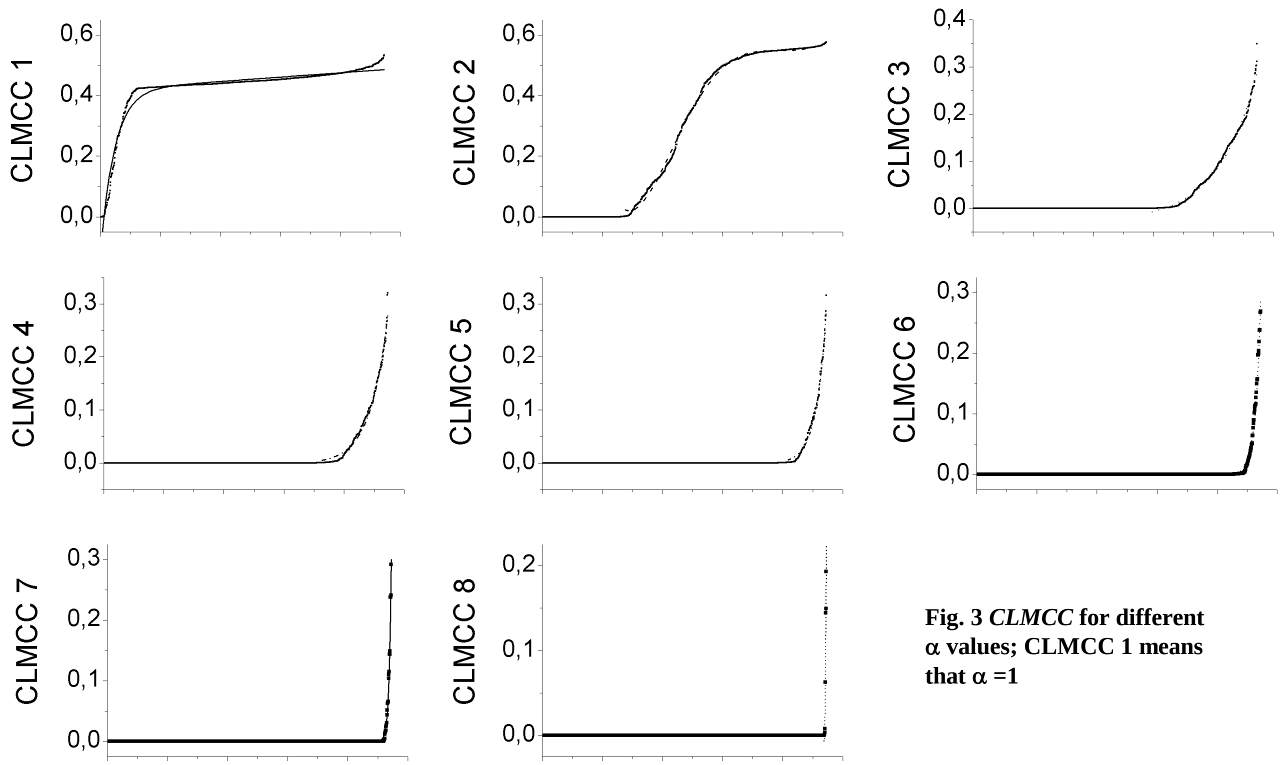

Fig. 3 ***CLMCC*** **for different α values; CLMCC 1 means that α =1**

6. References

[1] Barrat A., Barthelemy M., Vespignani A. Dynamical Processes on Complex Networks, Cambridge University Press, UK, 2008.

[2] Bródka P., Musial K., Kazienko P.: A Method for Group Extraction in Complex Social Networks. WSKS 2010, The 3rd World Summit on the Knowledge Society, Corfu, Greece, September, 22-24, 2010

[3] Caldarelli, G., Vespignani, A. (eds.): Large Scale Structure and Dynamics of Complex Networks, From Information Technology to Finance and Natural Science, Complex Systems and Interdisciplinary Science, vol. 2, World Scientific Publishing Co. Pte. Ltd., 2007.

[4] Hanneman, R., Riddle, M.: Introduction to social network methods, online textbook, Riverside, CA: University of California, (2005), http://faculty.ucr.edu/~hanneman/nettext/

[5] Kazienko P., Musiał K., Kajdanowicz T.: Profile of the Social Network in Photo Sharing Systems. AMCIS 2008, Association for Information Systems (AIS), ISBN: 978-0-615-23693-3.

[6] Scott, J.: Social network analysis: A handbook (2nd ed.), London, Sage (2000).

[7] Strogatz S. H. Exploring complex networks, Nature, 410(6825), March 1998, 268–276.

[8] Wasserman, S., Faust, K.: Social network analysis: Methods and applications. Cambridge University Press, New York, 1994.

[9] Watts D. Small Worlds: Dynamic of Networks between Order and Randomness, Princeton University Press, 2002.
[10] Watts D.J., Strogatz S., Collective dynamics of 'small-world' networks. Nature 393, June 1998, 440–444.