Bridging Online and Offline Social Networks: Multiplex Analysis

Andrej Gajduk[1], Sonja Filiposka[2], Ljupco Kocarev[1,2,3,*]

1 Macedonian Academy of sciences and Arts, Skopje, Macedonia
2 Faculty of Computer Science and Engineering, "Ss Cyril and Methodius University", Skopje, Macedonia
3 BioCircuits Institute, UC San Diego, La Jolla, CA 92093-0402, USA

*** E-mail: lkocarev@ucsd.edu**

Abstract

Abstract

Introduction

Materials and Methods

Subsection 1

Two groups of undergraduate students from Cyril and Methodius University, Skopje, Macedonia participated in this study. Students within one group are classmates attending during a semester the same class.

The total number of the first group, Group1, students (classmates) is ??, out of which (kolku studenti imaat razlicen od nula out degree) participated in the study. The ages of the ? female and ? the male students who participated ranged from 18 years to 22 years (M ?). There were ? juniors, ? seniors, ? sophomores, and ? freshmen (dali gi imame ovi brojki ili ne se vazni).

The total number of the second group, Group2, students is, while only ?? participated in the study. to add basic statistics.

Different data were also collected for each student through the university online electronic system (known as iknow system) (what data is available, what should be mentioned - not all of them)

The online survey was developed by the research team. Each student (who participated in the study) was asked to select from the list of classmates those with whom she/he was engaged in face-to-face communication and Facebook communication. The following categories for frequencies of contacts are used in this study. For the face-to-face communication: strong contact (tie) = more than 5 times per month; weak contact (tie) = more than once in 3 months, less than five times per month. For the Facebook communication: strong contact (tie) = more than five times per week; weak contact (tie) = more than once in a month, but less than five times per week.

Subsection 2

Graphs provide a powerful primitive for modeling data in social science. Nodes usually represent real world objects and edges indicate relationships between objects. In sociology, nodes may have attributes associated with them and graphs may contain many different types of relationships. The node attributes are used to describe the features of the objects that the nodes represent. For example, a node representing a student may have attributes that represent the student's gender and department. Different types of edges in a graph correspond to different types of relationships between nodes, such as friends and classmates relationships.

Here we study friendship relations among n actors: the existence of a tie $i \to j$ will be described as i calling j a friend. The ties are represented as binary variables, denoted by x_{ij}. A tie from actor i to actor j, $i \to j$, is either present or absent (x_{ij} then having values 1 and 0, respectively). The tie variables constitute the network, represented by its $n \times n$ adjacency matrix $X = [x_{ij}]$ (self-ties are excluded). The graph is directed, where each tie $i \to j$ has a sender i, who will also be referred to, which is common in social network analysis, as ego, and a receiver j, referred to as alter.

We model different types of relationships among actors with the concept of multiplex graphs. In sociology, multiplex graphs (networks) refer to the case when nodes (actors) are connected through more than one type of (socially relevant) tie. In mathematics, such graphs are also called multigraphs (a multigraph is a graph which is permitted to have multiple edges, that is, edges that have the same end nodes). We give a mathematical definition of multiplex graphs adapted for the study reported here. Let V denote the set of nodes; nodes are connected through L different type of connections (ties). Each type of connection (together with node set) forms a (directed) graph with n vertices. We denote with $G^\alpha = (V, E^\alpha)$ the graph which represents the connection α, where E^α denotes the set of type-α ties, $\alpha = 1, \ldots, L$. Let $[x_{ij}(G^\alpha)]$ be $n \times n$ adjacency matrix of the graph G^α. A multiplex graph $\mathcal{G}$ is then defined as a collection of all graphs G^α and all edge-aggregated graphs of the form $(V, \cup_{\alpha=\alpha_1}^{\alpha_2} E^\alpha)$, where $\alpha_1, \alpha_2 \in \{1, \ldots, L\}$. We assume that the case $\alpha_1 = \alpha_2$ is not excluded and write

$$G^{\alpha_1 \ldots \alpha_2} = (V, E^{\alpha_1 \ldots \alpha_2})$$

for the multiplex graph $\mathcal{G}$.

In the study discussed here, different types of edges correspond to friendship relations: online weak or strong connections and offline weak or strong connections. In addition, nodes in a graph have a set of associated attributes, which is denoted as $\{a_1, a_2, \ldots, a_t\}$.

Network characteristics

We consider a number of social network characteristics (parameters); those depending only on the network are called structural or endogenous characteristics, while parameters depending also on externally given attributes are called covariate or exogenous characteristics. Unifying framework for discussing all graph characteristics is the Jaccard index also known as Jaccard similarity coefficient. It measures similarity between finite sets and is defined as the size of the intersection divided by the size of the union of the sets:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|},$$

where $|S|$ denotes the cardinality of the set S. If A and B are both empty, we define $J(A, B) = 1$. Note that

$$0 \le J(A, B) \le 1.$$

The Jaccard distance, which measures dissimilarity between sample sets and is a metric on the collection of all finite sets, is complementary to the Jaccard coefficient and is obtained by subtracting the Jaccard coefficient from 1.

Let $G = (V, E)$ be a directed graph and let $S_i^{out}(G)$ and S_i^{in}(G) be a set of outgoing and ingoing neighbors of the node i, respectively, defined as:

$$S_i^{out}(G) = \{j : i \to j \in E\}, \quad S_i^{in}(G) = \{j : j \to i \in E\}.$$

The most basic characteristics of the graph G are out-degree and in-degee of the actor i, defined as

$$\begin{aligned} d_i^{out}(G) &= \sum_j x_{ij}(G) = \left|S_i^{out}(G)\right|, \\ d_i^{in}(G) &= \sum_j x_{ji}(G) = \left|S_i^{in}(G)\right|. \end{aligned}$$

Overlap

For a multiplex graph for which α and β represent two types of social relationships, we study overlapping index at the node i defined as

$$\begin{aligned} o_i^{out}(G^\alpha, G^\beta) &= J\left(S_i^{out}(G^\alpha), S_i^{out}(G^\beta)\right) \\ &= \frac{\left|S_i^{out}(G^\alpha) \cap S_i^{out}(G^\beta)\right|}{\left|S_i^{out}(G^\alpha) \cup S_i^{out}(G^\beta)\right|} \\ &= \frac{\sum_j x_{ij}(G^\alpha) x_{ij}(G^\beta)}{d_i^{out}(G^\alpha) + d_i^{out}(G^\beta) - \sum_j x_{ij}(G^\alpha) x_{ij}(G^\beta)} \end{aligned}$$

and its global counterpart

$$o(G^\alpha, G^\beta) = J(E^\alpha, E^\beta) = \frac{\left|E^\alpha \cap E^\beta\right|}{\left|E^\alpha \cup E^\beta\right|},$$

where $G^\alpha = (V, E^\alpha)$ and $G^\beta = (V, E^\beta)$.

na slicen nacin moze da se definira $o_i^{in}(G^\alpha, G^\beta)$

Reciprocity

Let $G = (V, E)$ be a directed graph. Another quite basic characteristic is normalized reciprocity represented by the number of reciprocated ties of actor i and defined as

$$r_i(G) = J\left(S_i^{out}(G), S_i^{in}(G)\right) = \frac{\sum_j x_{ij}(G) x_{ji}(G)}{d_i^{out}(G) + d_i^{in}(G) - \sum_j x_{ij}(G) x_{ji}(G)}.$$

The quantity reciprocity can also be extended for multiplex graphs as

$$r_i(G^\alpha, G^\beta) = J\left(S_i^{out}(G^\alpha), S_i^{in}(G^\beta)\right)$$

Transitivity

Next to reciprocity, an essential feature in most social networks is transitivity, or transitive closure (sometimes called clustering): friends of friends become friends, or in graph-theoretic terminology: two-paths tend to be, or to become, closed. Two characteristics will be considered here: transitive triplets and three-cycles, defined respectively as

$$\begin{aligned}\text{TransTrip}_i(G) &= \sum_{j,h} x_{ih}(G)x_{ij}(G)x_{jh}(G)\\ \text{ThreeCycles}_i &= \sum_{j,h} x_{ij}(G)x_{jh}(G)x_{hi}(G)\end{aligned}$$

The transitive triplets effect measures transitivity for an actor i by counting the number of pairs j, h such that there is the transitive triplet structure. The number of three-cycles that actor i is involved in is another characteristic found in many social networks. The transitive triplets and the three-cycle characteristics both represent closed structures, but whereas the former is in line with a hierarchical ordering, the latter goes against such an ordering.

To define normalized transitivity metrics, we first define a set of the neighbors of the neighbors of the node i as

$$W_i(G) = \{h : i \to j \text{ and } j \to h, \text{ for all } j, h \in V\}$$

Now we have

$$\begin{aligned}tt_i(G) &= J(W_i(G), S_i^{out}(G)) = \frac{|W_i \cap S_i^{out}|}{|W_i \cup S_i^{out}|}\\ &\frac{\sum_{j,h} x_{ih}(G)x_{ij}(G)x_{jh}(G)}{d_i^{out} + |W_i| - \sum_{j,h} x_{ih}(G)x_{ij}(G)x_{jh}(G)}\end{aligned}$$

In a similar fashion

$$tc_i(G) = J(W_i(G), S_i^{in}(G))$$

Da se prosirat za multiplex graphs.

Balance

A characteristic closely related to transitivity is balance defined as

$$\text{Bal}_i = \frac{1}{n-2}\sum_{j=1}^{n} x_{ij} \sum_{h=1; h\neq i,j}^{n} (b_0 - |x_{ih} - x_{jh}|),$$

where b_0 is the mean of $|x_{ih} - x_{jh}|$. Balance is the tendency to have and create ties to other actors who make the same choices as ego. The extent to which two actors make the same choices can be expressed simply as the number of outgoing choices and non-choices that they have in common.

Popularity and Activity

ne stignav ova da go napisam, ama po mene ima 4 mnozestva od tipot

$$\begin{aligned}W_i^1(G) &= \{h : i \to j \text{ and } j \to h, \text{ for all } j, h \in V\}\\ W_i^2(G) &= \{h : i \to j \text{ and } h \to j, \text{ for all } j, h \in V\}\\ W_i^3(G) &= \{h : j \to i \text{ and } j \to h, \text{ for all } j, h \in V\}\\ W_i^4(G) &= \{h : j \to i \text{ and } h \to j, \text{ for all } j, h \in V\}\end{aligned}$$

i Pop and Act se cardinatlity of those 4 sets.

In-degree popularity of an actor i relates how her/his friends are considered as friends by others. On the other hand, out-degree popularity of an actor i relates to the friends of her/his friends. Since

$$\begin{aligned}\sum_j x_{ij}(G)\sqrt{\sum_k x_{kj}(G)} &\leq \sum_j x_{ij}(G)\sqrt{\mathrm{MaxIn}(G)}\\ &= \mathrm{Out}_i(G)\sqrt{\mathrm{MaxIn}(G)}\\ &\leq \mathrm{MaxOut}(G)\sqrt{\mathrm{MaxIn}(G)}\end{aligned}$$

normalized popularity based characteristics can be defined as:

$$\begin{aligned}\mathrm{PopIn}_i(G) &= \frac{\sum_j x_{ij}(G)\sqrt{\sum_k x_{kj}(G)}}{\mathrm{MaxOut}(G)\sqrt{\mathrm{MaxIn}(G)}}\\ \mathrm{PopOut}_i(G) &= \frac{\sum_j x_{ij(G)}\sqrt{\sum_k x_{jk}(G)}}{\mathrm{MaxOut}(G)\sqrt{\mathrm{MaxOut}(G)}}\end{aligned}$$

A close to 1 in-degree-related popularity characteristic implies that high in-degrees reinforce themselves, which will lead to a relatively high dispersion of the in-degrees (so called a Matthew effect in popularity as measured by in-degrees). A close to 1 out-degree-related popularity characteristic increases the association between in-degrees and out-degrees. For a multiplex graph for which α and β represent two types of social relationships, normalized in-degree popularity is defined as:

$$\mathrm{PopIn}_i(G^\alpha, G^\beta) = \frac{\sum_j x_{ij}(G^\alpha)\sqrt{\sum_k x_{kj}(G^\beta)}}{\mathrm{MaxOut}(G^\alpha)\sqrt{\mathrm{MaxIn}(G^\beta)}}$$

and describes how for the actor i her/his type-α friends are considered by other actors as type-β friends.

Da se spomne i za activity

$$\begin{aligned}\mathrm{ActIn}_i &= \sum_k x_{ik}\sqrt{\sum_k x_{ki}}\\ \mathrm{ActOut}_i &= \sum_k x_{ik}\sqrt{\sum_k x_{ik}}\end{aligned}$$

Exogenous actor covariates

Nodes in a graph may have a set of associated attributes, also called exogenous actor covariates. This set is denoted as $\{a_1, a_2, \ldots, a_t\}$. There are three basic characteristics: the ego characteristic, measuring whether actors with higher a_i values tend to nominate more friends and hence have a higher out-degree (which also can be called covariate-related activity effect or sender effect); the alter characteristic, measuring whether actors with higher a_i values will tend to be nominated by more others and hence have higher in-degrees (covariate-related popularity effect, receiver effect); and the similarity effect, measuring whether ties tend to occur more often between actors with similar values on a_i (homophily effect). These characteristics are defined as:

$$\begin{aligned}\text{V-ego}_i &= \sum_j x_{ij} a_i \\ \text{V-alter}_i &= \sum_j x_{ij} a_j \\ \text{V-similarity}_i &= \sum_j x_{ij}(\text{sim}_{ij} - \overline{\text{sim}})\end{aligned}$$

where $\text{sim}_{ij} = (1 - |a_i - a_j|/\Delta)$, with $\Delta = \max_{ij} |a_i - a_j|$. Two other characterizes are

$$\begin{aligned}\text{V-same}_i &= \sum_j x_{ij} I(a_i = a_j) \\ \text{V-ego-alter}_i &= \sum_j x_{ij} a_i a_j\end{aligned}$$

where $I(a_i = a_j) = 1$ if $a_i = a_j$, and 0 otherwise.

Results

Let V be the set of all students from a class (group). For each class, 4 directed graphs have been generated from the survey: offline/online social network with strong ties and offline/online social network with weak ties, defined as:

$$\begin{aligned}G_s^{of} &= (V, E_s^{of}), \quad E_s^{of} = \{i \to j \text{ is an offline strong tie}\} \\ G_w^{of} &= (V, E_w^{of}), \quad E_w^{of} = \{i \to j \text{ is an offline weak tie}\} \\ G_s^{on} &= (V, E_s^{on}), \quad E_s^{on} = \{i \to j \text{ is an online strong tie}\} \\ G_w^{on} &= (V, E_w^{on}), \quad E_w^{on} = \{i \to j \text{ is an online weak tie}\}\end{aligned}$$

From these 4 (basic) graphs, 5 more aggregated graphs are constructed:

$$\begin{aligned}G^{of} &= (V, E^{of}), \quad E^{of} = E_s^{of} \cup E_w^{of} \\ G^{on} &= (V, E^{on}), \quad E^{on} = E_s^{on} \cup E_w^{on} \\ G_s &= (V, E_s), \quad E_s = E_s^{on} \cup E_s^{of} \\ G_w &= (V, E_w), \quad E_w = E_w^{on} \cup E_w^{of} \\ G &= (V, E) \quad E = E^{of} \cup E^{on} = E_s \cup E_w\end{aligned}$$

What is the structure of each of these graphs? Compute (for each graph) the number of isolated nodes, the number of nodes with ingoing links only, and the number of nodes for which the number of outgoing links is different than 0 (this is the number of students which participated in the study). Compute (for each graph) the number of strongly connected components and the number of weakly connected components. Also compute for the largest SCC (strongly connected component): number of nodes, number of links, average path, and some other basic metrics.

Research hypotheses and questions which will be addressed in this study are broken down into 3 (or some other numbers - we will see) main topics.

(1) Hypothesis (Granovetter): Consider two arbitrary selected individuals A and B and the set of all persons with ties to either or both of them. The first hypothesis is: the stronger the tie between A and B, the larger the proportion of individuals in S to whom they will both be tied (connected by a weak or strong tie).

(2) Hypothesis: Online SN supports pre-existing (offline) social relations.

Ovoj broj (normaliziran) ni kazuva: kolku od offline studentite sto i gi smeta za prijateli (strong or weak) nego go smetaat za prijatel. Kakva e raspredelbata na ovie broevi vo mrezata? srednata vrednost? slicni zaklucosi moze da se donesat i za drugite mrezi. na primer moze da se ragleduva sledniot odnos

$$\frac{\sum_j x_{ij}(G_s)x_{ji}(G_w)}{\mathrm{Out}_i(G_s) - |S_i^s|}$$

so drugi zborovi i go smeta j za strong prijatel, ama j go smeta i za weak. dali ima takvi vrski voopsto? sto toa znaci. mozni se i drugi kombinacii.

Acknowledgments

References

1. Tom A.B. Snijders, Gerhard G. van de Bunt, Christian E.G. Steglich, "Introduction to stochastic actor-based models for network dynamics," Social Networks 32 (2010) 44-60