

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Multiple-Membership Communities Detection and Its Applications for Mobile Networks

Nikolai Nefedov
Nokia Research Center
ISI Lab, Swiss Federal Institute of Technology Zurich (ETHZ)
Switzerland

1. Introduction

The recent progress in wireless technology and growing spread of smart phones equipped with various sensors make it possible to record real-world rich-content data and compliment it with on-line processing. Depending on the application, mobile data processing could help people to enrich their social interactions and improve environmental and personal health awareness. At the same time, mobile sensing data could help service providers to understand better human behavior and its dynamics, identify complex patterns of users' mobility, and to develop various service-centric and user-centric mobile applications and services on-demand. One of the first steps in analysis of rich-content mobile datasets is to find an underlying structure of users' interactions and its dynamics by clustering data according to some similarity measures.

Classification and clustering (finding groups of similar elements in data) are well-known problems which arise in many fields of sciences, e.g., (Albert & Barabási, 2002; Flake et al, 2002; Wasserman & Faust, 1994). In cases when objects are characterized by vectors of attributes, a number of efficient algorithms to find groups of similar objects based on a metric between the attribute vectors are developed. On the other hand, if data are given in the relational format (causality or dependency relations), e.g., as a network consisting of N nodes and E edges representing some relation between the nodes, then the problem of finding similar elements corresponds to detection of communities, i.e., groups of nodes which are interconnected more densely among themselves than with the rest of the network.

The growing interest to the problem of community detection was triggered by the introduction of a new clustering measure called modularity (Girvan & Newman, 2002; 2004). The modularity maximization is known as the NP-problem and currently a number of different sub-optimal algorithms are proposed, e.g., see (Fortunato, 2011) and references within. However, most of these methods address network partitions into disjoint communities.

On the other hand, in practice communities are often overlapping. It is especially visible in social networks, where only limited information is available and people are affiliated to different groups, depending on professional activities, family status, hobbies, and etc. Furthermore, social interactions are reflected in multiple dimensions, such as users activities, local proximities, geo-locations and etc. These multi-dimensional traces may be presented as multi-layer graphs. It raises the problem of overlapping communities detection at different

hierarchical levels at single and multi-layer graphs.

In this chapter we present a framework for multi-membership communities detection in dynamical multi-layer graphs and its applications for missing (or hidden) link predictions/recommendations based on the network topology. In particular, we use modularity maximization with a fast greedy search (Newman, 2004) extended with a random walk approach (Lambiotte et al, 2009) to detect multi-resolution communities beyond and below the resolution provided by *max*-modularity. We generalize a random walk approach to a coupled dynamic systems (Arenas et al, 2006) and then extend it with dynamical links update to make predictions beyond the given topology. In particular, we introduce attractive and repulsive coupling that allow us to detect and predict cooperative and competitive behavior in evolving social networks.

To deal with overlapping communities we introduce a soft community detection and outline its possible applications in single and multi-layer graphs. In particular, we propose friend-recommendations in social networks, where new link recommendations are made as intra- and inter-clique communities completion and recommendations are prioritized according to topologically-based similarity measures (Liben-Nowel & Kleinberg, 2003) modified to include multiple-communities membership. We also show that the proposed prediction rules based on soft community detection are in line with the network evolution predicted by coupled dynamical systems. To test the proposed framework we use a benchmark network (Zachary, 1977) and then apply developed methods for analysis of multi-layers graphs built from real-world mobile datasets (Kiukkonen et al, 2010). The presented results show that by combining information from multi-layer graphs we can improve reliability measures of community detection and missing links predictions.

The chapter is organized as follows: in Section 2 we outline the dynamical formulation of community detection that forms the basis for the rest of the paper. Topology detection using coupled dynamical systems and its extensions to model a network evolution are described in Section 3. Soft community detection for networks with overlapping communities and its applications are addressed in Section 4, followed by combining multi-layer graphs in Section 5. Evaluation of the proposed methods in the benchmark network are presented in Section 6. Analysis of some real-world datasets collected during Nokia data collection campaign is presented in Section 7, followed by conclusions in Section 8.

2. Community detection

2.1 Modularity maximization

Let's consider the clustering problem for an undirected graph $G = (V, E)$ with $|V| = N$ nodes and E edges. Recently Newman et al (Girvan & Newman, 2002; 2004) introduced a new measure for graph clustering, named a modularity, which is defined as a number connections within a group compared to the expected number of such connections in an equivalent null model (e.g., in an equivalent random graph). In particular, the modularity Q of a partition \mathcal{P} may be written as

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - P_{ij}) \delta(c_i, c_j), \quad (1)$$

where c_i is the i -th community, A_{ij} are elements of graph adjacency matrix; d_i is the i -th node degree, $d_i = \sum_j A_{ij}$; m is a total number of links $m = \sum_i d_i / 2$; P_{ij} is a probability that nodes i and j in a null model are connected; if a random graph is taken as the null model, then

$$P_{ij} = d_i d_j / 2m.$$

By construction $|Q| < 1$ and $Q = 0$ means that the network under study is equivalent to the used null model (an equivalent random graph). Case $Q > 0$ indicates a presence of a community structure, i.e., more links remain within communities than would be expected in an equivalent random graph. Hence, a network partition which maximizes modularity may be used to locate communities. This maximization is NP-hard and many suboptimal algorithms are suggested, e.g., see (Fortunato, 2011) and references within.

In the following we use the basic greedy search algorithm (Newman, 2004) extended with a random walk approach, since it gives a reasonable trade-off between accuracy of community detection and scalability.

Greedy Search Algorithm

Input: a weighted graph described by $N \times N$ adjacency matrix \mathbf{A} .

Initialize each node i as a community c_i with modularity $Q(i) = - \left(\frac{d_i}{2m} \right)^2$.

Repeat until there is an increase in modularity:

for all nodes i **do**:

- remove i from its community c_i ;
- insert i sequentially in neighboring communities c_j for all j with $A_{ij} > 0$;
- calculate modularity $Q(c_j^{(i \rightarrow c_j)})$;
- select a move (if any) of i -th node to community c_j^* with max modularity

$$Q(c_j^{(i \rightarrow c_j^*)}) = \max_{j \in N(i)} Q(c_j^{(i \rightarrow c_j)});$$

Stop when (a local) maximum is reached.

2.2 Communities detection with random walk

It is well-known that a network topology affects a system dynamics, it allows us to use the system dynamics to identify the underlying topology (Arenas et al, 2006; 2008; Lambiotte et al, 2009). First, we review the Laplacian dynamics formalism recently developed in (Evans & Lambiotte, 2009; Lambiotte et al, 2009).

Let's consider N independent identical Poisson processes defined on every node of a graph $G(V, E), |V| = N$, where random walkers are jumping at a constant rate from each of the nodes. We define p_n as the density of random walkers on node i at step n , then its dynamics is given by

$$p_{i,n+1} = \sum_j \frac{A_{ij}}{d_j} p_{j,n}. \quad (2)$$

The corresponding continuous-time process, described by (3),

$$\frac{dp_i}{dt} = \sum_j \frac{A_{ij}}{d_j} p_j - p_i = \sum_j \left(\frac{A_{ij}}{d_j} - \delta_{ij} \right) p_j \quad (3)$$

is driven by the random walk operator $\frac{A_{ij}}{d_j} - \delta_{ij}$, which in case of a discrete time process is presented by the random walk matrix $\mathbf{L}_{rw} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is a Laplacian matrix, \mathbf{A} is a non-negative weighted adjacency matrix, $\mathbf{D} = \text{diag}\{d_i\}$, $i = 1, \dots, N$. For an undirected connected network the stationary solution of (2) is given by $p_i^* = d_i/2m$. Let's now assume that for an undirected network there exist a partition \mathcal{P} with communities $c_k \in \mathcal{P}$, $k = 1, \dots, N_c$. The probability that initially, at t_0 , a random walker belongs to a community c_k is $\Pr(c_k, t_0) = \sum_{j \in c_k} d_j/2m$. Probability that a random walker, which was initially in c_k , will stay in the same community at the next step $t_0 + 1$ is given by

$$\Pr(c_k, t_0, t_0 + 1) = \sum_{j \in c_k} \sum_{i \in c_k} \left(\frac{A_{ij}}{d_j} \right) \left(\frac{d_j}{2m} \right). \quad (4)$$

The assumption that dynamics is ergodic means that the memory of the initial conditions are lost at infinity, hence $\Pr(c_k, t_0, \infty)$ is equal to the probability that two independent walkers are in c_k ,

$$\Pr(c_k, t_0, \infty) = \left(\sum_{i \in c_k} \frac{d_i}{2m} \right) \left(\sum_{j \in c_k} \frac{d_j}{2m} \right). \quad (5)$$

Combining (4) and (5) we may write

$$\sum_{c_k \in \mathcal{P}} (\Pr(c_k, t_0, t_0 + 1) - \Pr(c_k, t_0, \infty)) = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_i, c_j) = Q. \quad (6)$$

In general case, using (3), one may define a stability of the partition \mathcal{P} as (Evans & Lambiotte, 2009; Lambiotte et al, 2009)

$$R_{\mathcal{P}}(t) = \sum_{c_k \in \mathcal{P}} \Pr(c_k, t_0, t_0 + t) - \Pr(c_k, t_0, \infty) \quad (7)$$

$$= \sum_{c_k \in \mathcal{P}} \sum_{i,j \in c_k} \left(\left(e^{t(\hat{A}-I)} \right)_{ij} \frac{d_j}{2m} - \frac{d_i d_j}{4m^2} \right), \text{ where } \hat{A}_{ij} = \frac{A_{ij}}{d_j}. \quad (8)$$

Then, as the special cases of (8) at $t = 1$, we get the expression for modularity (6). Note that $R_{\mathcal{P}}(t)$ is non-increasing function of time: at $t = 0$ we get

$$R_{\mathcal{P}}(0) = 1 - \sum_{c_k \in \mathcal{P}} \sum_{i,j \in c_k} \frac{d_i d_j}{4m^2} \quad (9)$$

and $\max_{\mathcal{P}} R(0)$ is reached when each node is assigned to its own community. Note that (9) corresponds to collision entropy or Rényi entropy of order 2.

On the other hand, in the limit $t \rightarrow \infty$, the maximum of $R_{\mathcal{P}}(t)$ is achieved with Fiedler spectral decomposition into 2 communities. In other words, time here may be seen as a resolution parameter: with time t increasing, the $\max_{\mathcal{P}} R(t)$ results in a sequence of hierarchical

partitions $\{\mathcal{P}_t\}$ with the decreasing numbers of communities.

Furthermore, as shown in (Evans & Lambiotte, 2009), we may define a time-varying modularity $Q(t)$ by linear terms in time expansion for $R(t)$ at $t \approx 0$,

$$R(t) \approx (1 - t) \cdot R(0) + t \cdot Q = Q(t), \quad (10)$$

which after substitution (6) and (9) gives

$$Q(t) = (1 - t) + \sum_{c_k \in \mathcal{P}} \sum_{i, j \in c_k} \left(\frac{A_{ij}}{2m} t - \frac{d_i d_j}{4m^2} \right). \quad (11)$$

In the following we apply time-dependent modularity maximization (11) using the greedy search to find hierarchical structures in networks beyond modularity maximization Q_{max} in (1). This approach is useful in cases where maximization of (1) results in a very fragmental structure with a large number of communities. Also it allows us to evaluate the stability of communities at different resolution levels. However, since the adjacency matrix \mathbf{A} is not time dependent, the time-varying modularity (11) can not be used to make predictions beyond the given topology.

3. Topology detection using coupled dynamical systems

3.1 Laplacian formulation of network dynamics

Let's consider an undirected weighted graph $G = \{V, E\}$ with N nodes and E edges, where each node represents a local dynamical system and edges correspond to local coupling. Dynamics of N locally coupled dynamical systems on the graph G is described by

$$\dot{x}_i(t) = q_i(x_i(t)) + k_c \sum_{j=1}^N A_{ij} \psi(x_j(t) - x_i(t)), \quad (12)$$

where $q_i(x_i)$ describes a local dynamics of state x_i ; A_{ij} is a coupling strength between nodes i and j ; $\psi(\cdot)$ is a coupling function; k_c is a global coupling gain.

In case of weakly phase-coupled oscillators the dynamics of local states is described by Kuramoto model (Acebron et al, 2005; Kuramoto, 1975)

$$\dot{\theta}_i(t) = \omega_i + k_c \sum_{j=1}^N A_{ij} \sin[\theta_j(t) - \theta_i(t)]. \quad (13)$$

Linear approximation of coupling function $\sin(\theta) \simeq \theta$ in (13) results in the consensus model (Olfati-Saber et al, 2007)

$$\dot{\theta}_i(t) = k_c \sum_{j=1}^N A_{ij} [\theta_j(t) - \theta_i(t)], \quad (14)$$

which for a connectivity graph G may be written as

$$\dot{\Theta}(t) = -k_c \mathbf{L} \Theta(t), \quad (15)$$

where $\mathbf{L} = \mathbf{A} - \mathbf{D}$ is the Laplacian matrix of G . The solution of (15) in the form of normal modes $\omega_i(t)$ may be written as

$$\omega_i(t) = k_c \sum_{j=1}^N V_{ij} \theta_j(t) = k_c \omega_i(t_0) e^{-\lambda_i t}, \quad (16)$$

where $\lambda_1, \dots, \lambda_N$ are eigenvalues and \mathbf{V} is the matrix of eigenvectors of \mathbf{L} . Note that (16) describes a convergence speed to a consensus for each nodes. Let's order these equations according to the descending order of their eigenvalues. Then it is easy to see that nodes are approaching the consensus in a hierarchical way, revealing in the same time a hierarchy of communities in the given network G .

Note that (15) has the same form as (3), with the difference that the random walk process (3) is based on $\mathbf{L}_{rw} = \mathbf{D}^{-1} \mathbf{L}$. It allows us to consider random-walk-based communities detection in the previous section as a special case of coupled oscillators synchronization.

Similarly to (15), we may derive the Laplacian presentation for locally coupled oscillators (13). In particular, the connectivity of a graph may be described by the graph incidence ($N \times E$) matrix \mathbf{B} : $\{\mathbf{B}\}_{ij} = 1$ (or -1) if nodes j and i are connected, otherwise $\{\mathbf{B}\}_{ij} = 0$. In case of weighted graphs we use the weighted Laplacian defined as

$$\mathbf{L}_A \triangleq \mathbf{B} \mathbf{D}_A \mathbf{B}^T. \quad (17)$$

Based on (17) we can rewrite (13) as

$$\dot{\Theta}(t) = \mathbf{\Omega} - k_c \mathbf{B} \mathbf{D}_A \sin(\mathbf{B}^T \Theta(t)), \quad (18)$$

where vectors and matrices are defined as follows:

$\Theta(t) \triangleq [\theta_1(t), \dots, \theta_N(t)]^T$; $\mathbf{\Omega}(t) \triangleq [\omega_1(t), \dots, \omega_N(t)]^T$; $\mathbf{D}_A \triangleq \text{diag}\{a_1, \dots, a_E\}$, a_1, \dots, a_E are weights A_{ij} indexed from 1 to E .

In the following we use (18) to describe different coupling scenarios.

3.2 Dynamical structures with different coupling scenarios

Let's consider local correlations between instant phases of oscillators,

$$r_{ij}(t) = \langle \cos[\theta_j(t) - \theta_i(t)] \rangle, \quad (19)$$

where the average is taken over initial random phases $\theta_i(t=0)$.

Following (Arenas et al, 2006; 2008) we may define a dynamical connectivity matrix $\mathbf{C}_t(\eta)$, where two nodes i and j are connected at time t if their local phase correlation is above a given threshold η ,

$$\begin{aligned} C_t(\eta)_{ij} &= 1 & \text{if } r_{ij}(t) > \eta \\ C_t(\eta)_{ij} &= 0 & \text{if } r_{ij}(t) < \eta. \end{aligned} \quad (20)$$

We select communities resolution level (time t) using a random walk as in Section 2. Next, by changing the threshold η , we obtain a set of connectivity matrices $\mathbf{C}_t(\eta)$ which reveal dynamical topological structures for different correlation levels. Since the local correlations $r_{ij}(t)$ are continuous and monotonic functions in time, we may also fix η and express

dynamical connectivity matrix (20) in the form $\mathbf{C}_\eta(t)$ to present the evolution of connectivity in time for a fixed correlation threshold η . Using this approach we consider below several scenarios of networks evolution with dynamically changing coupling.

B.1. Attractive coupling with dynamical updates

As the first step, let's introduce dynamics into static attractive coupling (13). Using the dynamical connectivity matrix (20) we may write

$$\dot{\theta}_i(t) = \omega_i + k_c \sum_{j=1}^N F_{ij}^{(\eta)}(t) \sin [\theta_j(t) - \theta_i(t)], \quad (21)$$

where matrix $\mathbf{F}^{(\eta)}(t)$ describes dynamical attractive coupling, $F_{ij}^{(\eta)}(t) = A_{ij} \mathbf{C}_\eta(t)_{ij} \geq 0$. Then, similar to (18), the attractive coupling with a dynamical update may be described as

$$\dot{\Theta}(t) = \mathbf{\Omega} - k_c \mathbf{B}(t) \mathbf{D}_F(t) \sin (\mathbf{B}(t)^T \Theta(t)), \quad (22)$$

where initial conditions are defined by A_{ij} ; $\mathbf{D}_F(t)$ is formed from \mathbf{D}_A with elements $\{a_k\}$ scaled according to $\mathbf{C}_\eta(t)$.

B.2. Combination of attractive and repulsive coupling with dynamical links update

Many biological and social systems show a presence of a competition between conflicting processes. In case of coupled oscillators it may be modeled as the attractive coupling (driving oscillators into the global synchronization) combined with the repulsive coupling (forcing system into a chaotic/random behavior). To allow positive and negative interactions we use instant correlation matrix $\mathbf{R}(t) = \mathbf{R}^+(t) + \mathbf{R}^-(t)$, and separate attractive and repulsive parts

$$\dot{\theta}_i(t) = \omega_i + k_c \sum_{j=1}^N r_{ij}^+(t) A_{ij} \sin [\theta_j(t) - \theta_i(t)] - k_c \sum_{j=1}^N |r_{ij}^-(t)| A_{ij} \sin [\theta_j(t) - \theta_i(t)], \quad (23)$$

where superscripts denote positive and negative correlations¹.

Note that the total number of links in the network does not change, at a given time instant each link performs either attractive or repulsive function.

To obtain the Laplacian presentation we define a dynamical connectivity matrix $\mathbf{F}(t)$ as element-by-element matrix product

$$\mathbf{F}(t) = \mathbf{R}(t) \circ \mathbf{A} = \mathbf{F}^+(t) + \mathbf{F}^-(t), \quad (24)$$

and present dynamic Laplacian as the following

$$\mathbf{L}_F(t) = \mathbf{B}(t)(\mathbf{D}_{F^+}(t) + \mathbf{D}_{F^-}(t))\mathbf{B}^T(t). \quad (25)$$

It allows us to write

$$\dot{\theta}_i(t) = \omega_n + k_c \sum_{m=1}^N F_{ij}^+(t) \sin [\theta_j(t) - \theta_i(t)] - k_c \sum_{m=1}^N F_{ij}^-(t) \sin [\theta_j(t) - \theta_i(t)], \quad (26)$$

¹ For presentation clarity we omit here the correlation threshold η .

or in matrix form

$$\dot{\Theta}(t) = \Omega - k_c \mathbf{B}(t) \mathbf{D}_{F^+}(t) \sin(\mathbf{B}^T(t) \Theta(t)) + k_c \mathbf{B}(t) \mathbf{D}_{F^-}(t) \sin(\mathbf{B}^T(t) \Theta(t)). \quad (27)$$

B.3. Combination of attractive and initially neutral coupling with dynamical links update

Negative correlations (resulting in repulsive coupling) are typically assigned between nodes which are not initially connected. However, in many cases this scenario is not realistic. For example, in social networks, the absence of communications between people does not necessarily indicate conflicting (negative) relations, but often has a neutral meaning. To take this observation into account we modified second term in (23) such that it sets neutral initial conditions to unconnected nodes in adjacency matrix \mathbf{A} . In particular, system dynamics with links update (24) and initially neutral coupling is described by

$$\dot{\theta}_i(t) = \omega_i + k_c \sum_{j=1}^N F_{ij}^+(t) \sin[\theta_j(t) - \theta_i(t)] + k_c \sum_{j=1}^N F_{ij}^-(t) \cos[\theta_j(t) - \theta_i(t)], \quad (28)$$

or in the matrix form

$$\dot{\Theta}(t) = \Omega - k_c \mathbf{B}(t) \mathbf{D}_{F^+}(t) \sin(\mathbf{B}^T(t) \Theta(t)) - k_c \mathbf{B}(t) \mathbf{D}_{F^-}(t) \cos(\mathbf{B}^T(t) \Theta(t)). \quad (29)$$

Then a dynamical interplay between the given network topology and local interactions drives the connectivity evolution. We evaluated the scenarios above using different clustering measures (Manning et al, 2008) and found that scenario B.3 shows the best performance.

In the following we use coupled system dynamics approach to predict networks' evolution and to make missing links predictions and recommendations. Furthermore, the suggested approach allows us also to predict repulsive relations in the network based on the network topology and links dynamics.

4. Overlapping communities

4.1 Multi-membership

In social networks people belong to several overlapping communities depending on their families, occupations, hobbies, etc. As the result, users (presented by nodes in a graph) may have different levels of membership in different communities. This fact motivated us to consider multi-community membership as edge-weights to different communities and partition edges instead of clustering nodes.

As an example, we can measure a membership $g_j(k)$ of node k in j -th community as a number of links (or its weight for a weighted graph) between the k -th node and other nodes within the same community, $g_j(k) = \sum_{i \in c_j} w_{ki}$. Then, for each node k we assign a vector $\mathbf{g}(k) = [g_1(k), g_2(k), \dots, g_{N_c}(k)]$, $k \in \{1, \dots, N\}$ which presents the node membership (or participation) in all detected communities $\{c_1, \dots, c_{N_c}\}$. In the following we refer $\mathbf{g}(k)$ as a soft community decision for the k -th node.

To illustrate the approach, overlapping communities derived from benchmark karate club social network (Zachary, 1977) and membership distributions for selected nodes are depicted

at Fig.1 and Fig.2, respectively. Modularity maximization here reveals 4 communities shown by different colors. However, the multi-communities membership results in overlapping communities illustrated by overlapping ovals (Fig.1). For example, according to modality maximization, the node 1 belongs to community c_2 , but it also has links to all other communities indicated by blue bars at Fig.2.

Participation of different nodes in selected communities is presented at Fig.3 and Fig.4. These graphs show that even if a node is assigned by some community detection algorithm to a certain community, it still may have significant membership in other communities. This multi-communities membership is one of the reasons why different algorithms disagree on communities partitions. In practice, e.g., in targeted advertisements, due to the "hard" decision in community detection, some users may be missed even if they are strongly related to the targeted group. For example, user '29' is assigned to c_3 (Fig.1), but it also has equally strong memberships in c_2 and c_4 (Fig.3). Using soft community detection user '29' can also be qualified for advertisements targeted to c_2 or c_4 .

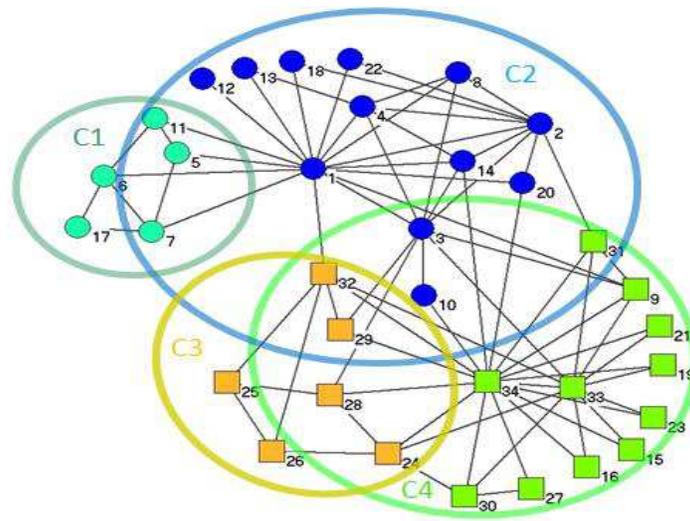


Fig. 1. Overlapping communities in karate club.

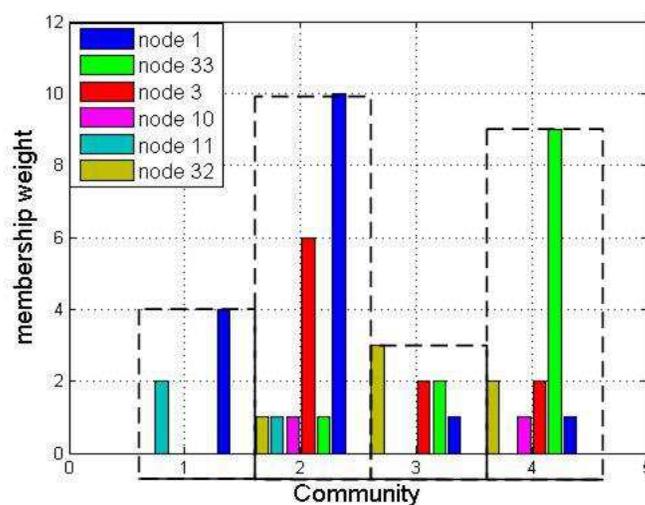


Fig. 2. Membership weight distribution for selected users in karate club social network.

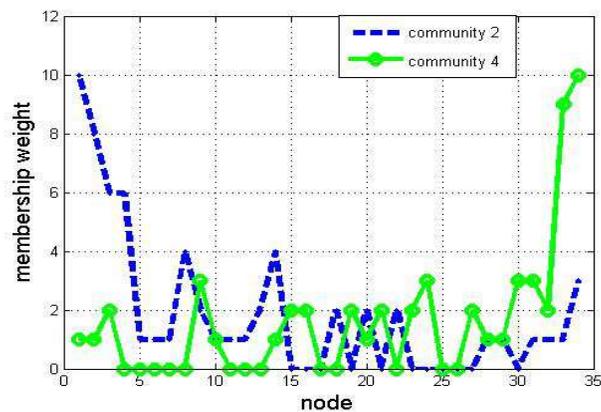


Fig. 3. Karate club: participation of users in communities c_2, c_4 .

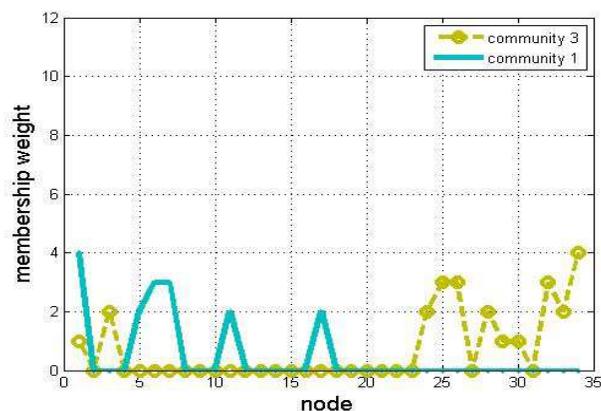


Fig. 4. Karate club: participation of users in communities c_1, c_3 .

4.2 Application of soft community detection for recommendation systems

In online social networks a recommendation of new social links may be seen as an attractive service. Recently Facebook and LinkedIn introduced a service "People You May Know", which recommends new connections using the friend-of-friend (FoF) approach. However, in large networks the FoF approach may create a long and often not relevant list of recommendations, which is difficult (and also computationally expensive, in particular in mobile solutions) to navigate. Furthermore, in mobile social networks (e.g., Nokia portal Ovi Store) these kinds of recommendations are even more complicated because users' affiliations to different groups (and even its number) are not known. Hence, before making recommendations, communities are to be detected first.

Recommendations as communities completion

Based on soft communities detection we suggest to make the FoF recommendations as follows:

- (i) detect communities, e.g., by using one of the methods described above;
- (ii) calculate membership $g_j(k)$ in all relevant communities for each node k ;
- (iii) make new recommendations as communities completion following the rules below;
- (iv) use multiple-membership to prioritize recommendations.

To make new link recommendations in (iii) we suggest the following rules:

- each new link creates at least one new clique (the FoF concept);
- complete cliques within the same community (intra-cliques) using the FoF concept;
- complete cliques towards to the fully-connected own community if there is no FoF links;
- complete inter-cliques (where nodes belong to different communities);
- prioritize intra-clique and inter-clique links completion according to some measure based on multi-membership.

To assign priorities we introduce several similarity measures outlined below. We will show in next sections that these rules are well in line with link predictions made by coupled dynamical systems described in Section 3.

Modified topology-based predictors

Let's define sets of neighbors of node k , which are inside and outside of community c_i as $\Gamma_i(k) = \{\Gamma(k) \in c_i\}$ and $\Gamma_{\setminus i}(k) = \{\Gamma(k) \notin c_i\}$, respectively. This allows us to introduce a set of similarity measures by modifying topology-based base-line predictors listed in (Liben-Nowel & Kleinberg, 2003) to take into account the multiple-membership in overlapping communities.

As an example, for the intra-clique completion we may associate a quality of missing link prediction (or recommendation) between nodes k and n within c_i community by modifying the base-line predictor scores as follows:

- Preferential attachment: $S_{PA}^{(i,i)}(k, n) = |\Gamma_i(k)| \cdot |\Gamma_i(n)|$;
- Jaccards score: $S_{JC}^{(i,i)}(k, n) = |\Gamma_i(k) \cap \Gamma_i(n)| / |\Gamma_i(k) \cup \Gamma_i(n)|$;
- Adamic/Adar score: $S_{AA}^{(i,i)}(k, n) = \sum_{z \in \Gamma_i(k) \cap \Gamma_i(n)} (\log |\Gamma(z)|)^{-1}$;
- Katz score (intra-community):

$$S_{KC}^{(i,i)}(k, n) = \sum_{l=1}^{\infty} \beta^l |\text{path}(k, n)^{(l)}| = \left\{ (I - \beta A^{(i)})^{-1} - I \right\}_{(k,n)},$$

where $|\text{path}_i(k, n)^{(l)}|$ is number of all paths of length- l from k to n within c_i ; I is the identity matrix, $A^{(i)}$ is the (weighted) adjacency matrix of community c_i , β is a dumping parameter, $0 < \beta < 1$, such that $\sum_{ij} \beta A_{ij} < 1$.

Additionally to the base-line predictors above, we also use a community connectivity measure, $S_{CC}^{(i,i)}(k, n) \sim \sigma_2(L_i)$, which characterizes a convergence speed of opinions to consensus within a community c_i when a link between nodes k and n is added inside the community; here $\sigma_2(L)$ is the 2nd smallest eigenvalue of the graph Laplacian L_i of community c_i (or the normalized Laplacian for weighted graphs, based on (17)).

The measures above consider communities as disjoint sets and may be used as the 1st order approximation for link predictions in overlapping communities. To take into account both intra- and inter-community links we use multi-community membership for nodes, $g_i(k)$. In general, for nodes $k \in c_i$ and $n \in c_j$, the inter-community relations may be asymmetric, $g_j(k) \neq g_i(n)$. In the case of undirected graphs we may use averaging and modify the base-line predictors $S(k, n)$ as

$$S^{(i,j)}(k, n) = \frac{g_j(k) + g_i(n)}{2m} S(k, n). \quad (30)$$

For example, modified Jaccards and Katz scores which take into account multi-communities membership are defined as

$$S_{JC}^{(i,j)}(k,n) = \frac{g_j(k) + g_i(n)}{2m} \frac{|\Gamma(k) \cap \Gamma(n)|}{|\Gamma(k) \cup \Gamma(n)|}, \quad (31)$$

$$S_{KC}^{(i,j)}(k,n) = \frac{g_j(k) + g_i(n)}{2m} \left\{ (I - \beta A^{(C_{n,k})})^{-1} - I \right\}_{(k,n)}, \quad (32)$$

where $k \in c_i, n \in c_j$; $A^{(C_{n,k})}$ is an adjacency matrix formed by all communities relevant to nodes n and k .

Recommendations also may be made in the probabilistic way, e.g., to be picked up from distributions formed by modified prediction scores.

5. Multi-layer graphs

In analysis of multi-layer graphs we assume that different network layers capture different modalities of the same underlying phenomena. For example, in case of mobile networks the social relations are partly reflected in different interaction layers, such as phone and SMS communications recorded in call-logs, people "closeness" extracted from the bluetooth (BT) and WLAN proximity, common GPS locations and traveling patterns and etc. It may be expected that a proper merging of data encoded in multi-graph layers can improve the classification accuracy.

One approach to analyze multi-layer graphs is first to merge graphs according to some rules and then extract communities from the combined graph. The layers may be combined directly or using some functions defined on the graphs. For example, multiple graphs may be aggregated in spectral domain using a joint block-matrix factorization or a regularization framework (Dong et al, 2011). Another method is to extract spectral structural properties from each layer separately and then to find a common presentation shared by all layers (Tang et al, 2009).

In this paper we consider methods of combining graphs based on modularity maximization

$$\max Q = \max_{c_i, c_j} \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(c_j, c_j). \quad (33)$$

Let's define a modularity matrix \mathbf{M} with elements $M_{ij} = A_{ij} - \frac{d_i d_j}{2m}$. Then the modularity in (33) may be presented as

$$Q = \frac{1}{2m} \text{Tr} \left(\mathbf{G}^T \left(\mathbf{A} - \frac{\mathbf{d}\mathbf{d}^T}{2m} \right) \mathbf{G} \right) = \frac{1}{2m} \text{Tr}(\mathbf{G}^T \mathbf{M} \mathbf{G}), \quad (34)$$

where columns of $N \times N_c$ matrix \mathbf{G} describes community memberships for nodes, $g_j(i) = g_{ij} \in \{0,1\}$, $g_{ij} = 1$ if the i -th node belongs to the community c_j ; N_c is a number of communities; \mathbf{d} is a vector formed by degrees of nodes, $\mathbf{d} = (d_1, \dots, d_N)^T$.

Let's consider a multi-layer graph $\mathcal{G} = \{G_1, G_2, \dots, G_L\}$ with adjacency matrices $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_L\}$, where L is a number of layers. Before combining, the graphs are to be normalized. In case of modularity maximization (33) it is natural to normalize each layer

according its total weight m .

The simplest method to combine multi-layer graphs is to make the average of all layers:

$$\bar{\mathbf{A}} = \frac{1}{L} \sum_l \mathbf{A}_l; \quad \bar{\mathbf{d}} = \frac{1}{L} \sum_l \mathbf{d}_l; \quad \bar{m} = \frac{1}{L} \sum_l m_l; \quad \max_Q = \max_{\mathbf{G}} \frac{1}{2\bar{m}} \text{Tr}(\mathbf{G}^T \bar{\mathbf{M}} \mathbf{G}) \quad (35)$$

Then the community membership matrix \mathbf{G} may be found by one of community detection methods described before. By taking into account degree distributions of nodes at each graph layer, the total modularity across all layers may maximized as (Tang et al, 2009)

$$\max_Q = \frac{1}{L} \sum_l Q_l = \max_{\mathbf{G}} \frac{1}{2L} \sum_l \text{Tr} \left(\mathbf{G}^T \left(\mathbf{A}_l - \frac{\mathbf{d}_l \mathbf{d}_l^T}{2m_l} \right) \mathbf{G} \right) = \max_{\mathbf{G}} \frac{1}{L} \sum_l \text{Tr} \left(\mathbf{G}^T \frac{\mathbf{M}_l}{2m_l} \mathbf{G} \right), \quad (36)$$

Similar approach, but applied to graph Laplacian spectra and extended with a regularization, is used in (Dong et al, 2011).

Typically networks describing social relations are often undersampled, noisy and contain different amount of information at each layer. As the result, a noisy or an observable part(s) in one of the layers after averaging in (35) and (36) may deteriorate the total accuracy. A possible solution for this problem is to apply weighted superposition of layers. In particular, the more informative the layer l is, the larger weight w_l it should be given. For example, we may weight the layer l according to its modularity Q_l , hence

$$\bar{\mathbf{A}}_w = \frac{1}{L} \sum_l w_l \mathbf{A}_l = \frac{1}{L} \sum_l Q_l \mathbf{A}_l; \quad (37)$$

Another method to improve the robustness of nodes classification in multi-layer graphs is to extract structural properties \mathbf{G}_l at each layer separately and then merge partitions (Strehl & Ghosh, 2002). The more advanced approach of processing of multi-dimensional data may be based on presenting multi-layer graphs as tensors and apply tensor decomposition algorithms (Kolda & Bader, 2009) to extract stable communities and make de-noising by lower-dimension tensor approximation. These methods are rather involved and will be considered elsewhere.

6. Simulation results for benchmark networks

To test algorithms described in the previous sections we use the karate club social network (Zachary, 1977). As mentioned before, to get different hierarchical levels beyond and below the resolution provided by *max*-modularity, we use the random walk approach. A number of detected communities in the karate club at different resolution levels is presented at Fig.5. As one can see, the *max*-modularity algorithm does not necessary result in the best partition stability. The most stable partition in case of the karate club corresponds to 2 communities (shown by squares and circles at Fig.1), which is in line with results reported by (Zachary, 1977).

Comparison of coupling scenarios *B.2* and *B.3* is presented at Fig.6 and Fig.7. Pair-wise correlations between oscillators at $t = 1$ for coupling scenarios *B.2* and *B.3* are depicted at Fig.6. Scenario *B.3* reveals clearly communities structure, while in case of *B.2* the negative coupling overwhelms the attractive coupling and forces the system into a chaotic behavior.

Dynamical connectivity matrices reordered by communities for the attractive-neural coupling *B.3* at $t = 1$ (on the left) and $t = 10$ (on the right) are depicted at Fig.7. In case *B.3* one can see (also cf. Fig.8) that number of connections with the attractive coupling is growing in time, while the strength of the repulsive connections is decreasing, which finally results in the global synchronization. For the scenario *B.2* there is a dynamical balance between attractive and repulsive coupling with small fluctuations around the mean (Fig.8). Note that even the averaged strength of the repulsive connections is less than the attractive coupling, the system dynamics shows a quasi-chaotic behavior.

Fig.9 shows the adjacency matrix for Zachary karate club (red circles), detected communities by pink squares, predicted links are shown by blue dots. As expected, the dynamical methods for links prediction tend to make more connections within the established communities first, followed by merging communities and creating highly overlapped partitions at the higher hierarchical levels (the upper part at Fig.9). In case of Katz predictor (32), by increasing the dumping parameter β we take into account the larger number of paths connecting nodes in the graph, which in turn results into the larger number of suggested links above a fixed threshold. Following the concept of dynamical connectivity matrix (20), the process of growing number of links may be seen as the hierarchical community formation predicted by (32) at different values of β . This process is illustrated at Fig.9, the bottom part. Note that in case of Katz predictor, the connected graph is also approaching the fully connected graph, but the network evolution may take a different trajectory compared to the coupled dynamical systems. In particular, at small values of t and β , the network evolution is similar for both cases (cf. Fig.9(b) and Fig.9(e)), but with the time the evolution trajectories may follow different paths (cf. Fig.9(c) and Fig.9(f)), which in turn results in different predictions.

Note that in all cases of the network evolution, we may prioritize the recommended links based on the soft communities detection (Katz predictor) or the threshold η (coupled dynamical systems). We address this issue below in Section 7.

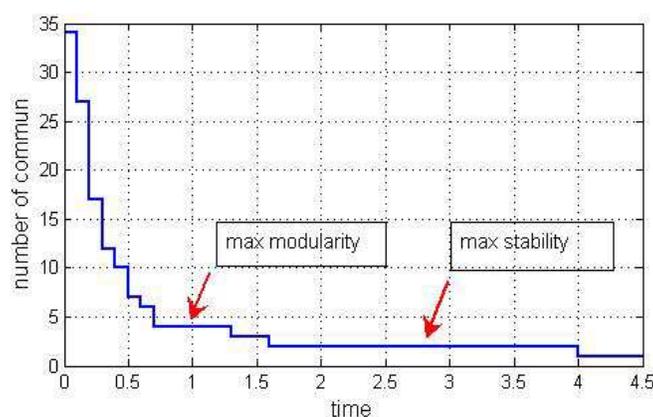


Fig. 5. Karate club: number of communities at different resolution levels.

7. Applications for real world mobile data

7.1 Community detection in Nokia mobile datasets

To analyze mobile users behavior and study underlying social structure, Nokia Research Center/Lausanne organized mobile data collection campaign at EPFL university campus (Kiukkonen et al, 2010). Rich-content datasets (including data from mobile sensors, call-logs,

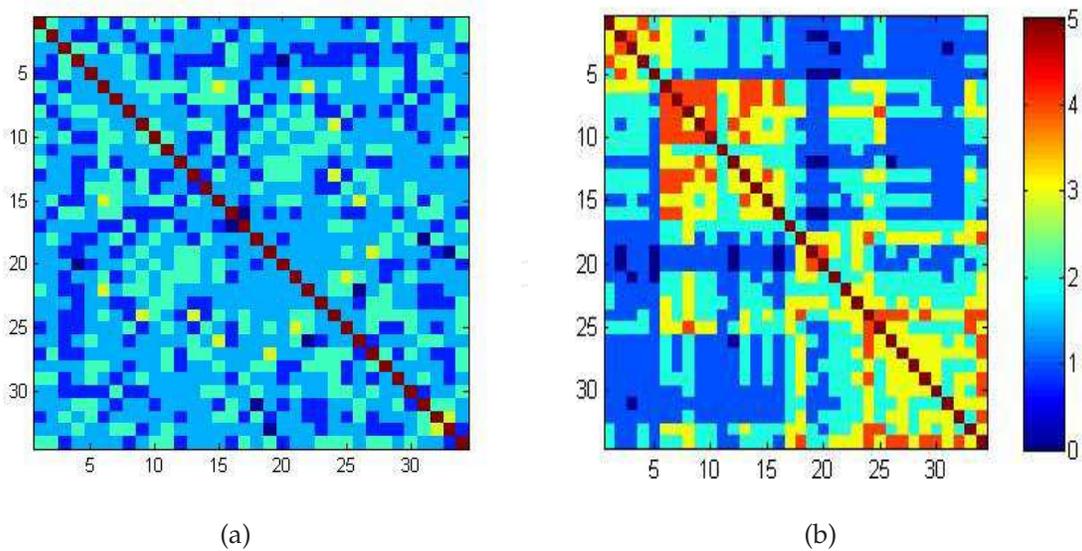


Fig. 6. Karate club: averaged pair-wise correlations (scaled by 5) between oscillators at $t = 1$ re-ordered according to communities. Coupling scenarios: (a) attractive-repulsive $B.2$; (b) attractive-neutral $B.3$.

bluetooth (BT) and WLAN proximity, GPS coordinates, information on mobile and applications usage and etc) are collected from about 200 participants for the period from June 2009 till October 2010. Besides the collected data, several surveys before and after the campaign have been conducted to profile participants and to form a basis for the ground truth. In this section we consider social affinity graphs constructed from call-logs, GPS locations and users proximity.

Fig.10 shows a weighted aggregated graph of voice-calls and SMS connections derived from corresponding datasets. This graph depicts connections among 136 users, which indicates that about 73% of participants are socially connected within the data collection campaign. To find communities in this network we first run the modularity maximization algorithm, which identifies 14 communities after the 3d iteration (Fig.10). To get the higher hierarchical levels one could represent each community by a single node and continue clustering with the new aggregated network. However, this procedure would result in a loss of underlying structure. In particular, the hierarchical community detection with the nested communities structure poses additional constrains on the maximization process and may lead to incorrect classification at the higher layers. For example, after the 3d iteration the node "v146", shown by red arrow at Fig.10, belongs (correctly) to a community shown by white circles (3 intra-community edges and single edges to other 6 communities). After agglomeration, the node "v146" will be assigned to the community shown by white circles on the left side of the graph. However, it is easy to verify that when communities on the right are merged, the node "v146" is to be re-assigned to the community on the right side of the network. Dynamical formulation of modularity extended with the random walk allows different (not necessarily nested) allocations of nodes at different granularity (resolution) levels and helps to resolve this problem.

Fig.11 presents a number of communities at different hierarchical levels detected by the random walk for the network shown at Fig.10. As one can see, the *max*-modularity partition with 14 communities is clearly unstable and hardly could be used for reliable predictions, the

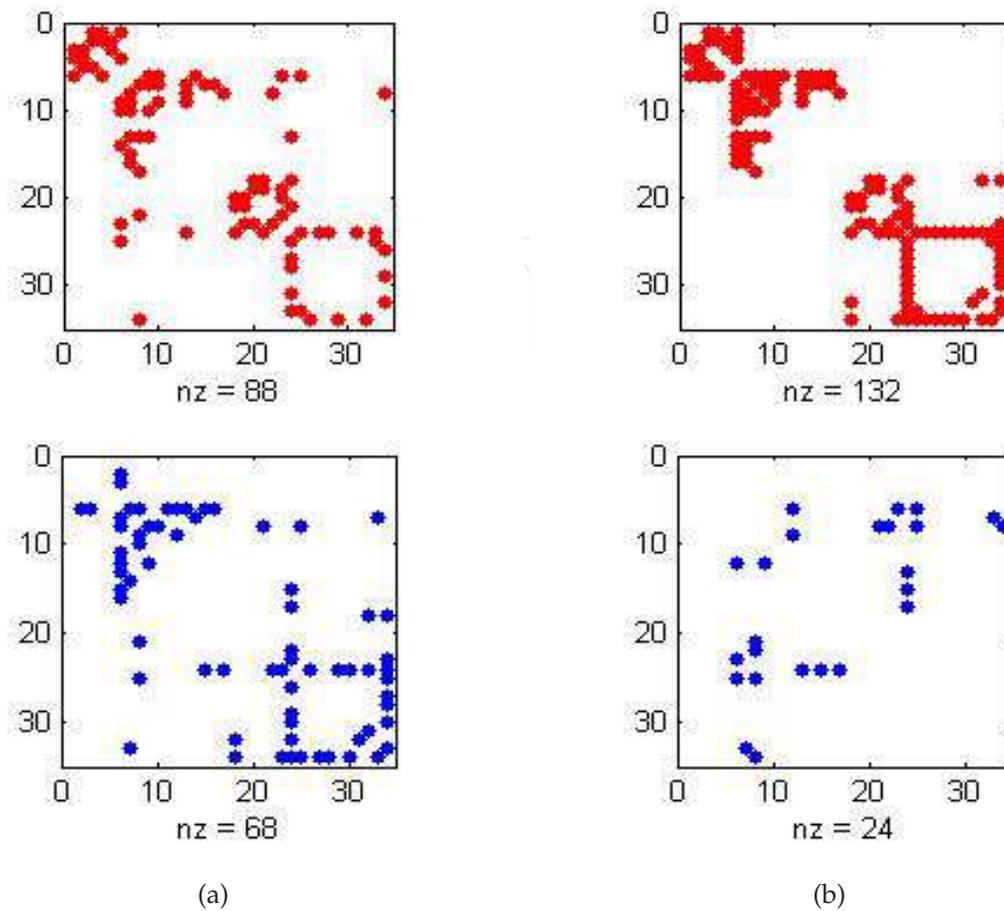


Fig. 7. Karate club: examples of dynamical connectivity matrices for attractive (shown on the top in red color) and repulsive (shown at the bottom in blue color) coupling at $t = 1$ (a) and $t = 10$ (b). Nodes are ordered according to communities. Coupling scenarios: attractive-neutral $B.3$.

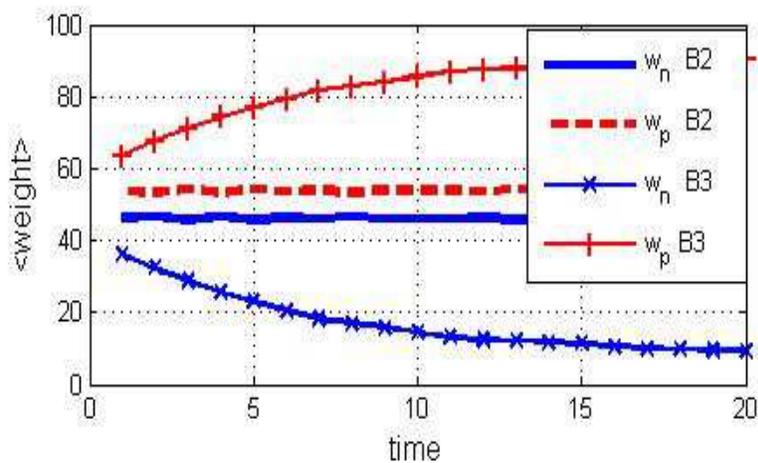


Fig. 8. Karate club: evolution of averaged attractive w_p and repulsive w_n weights for different coupling scenarios $B.2$ and $B.3$; the average is made over 100 realizations.

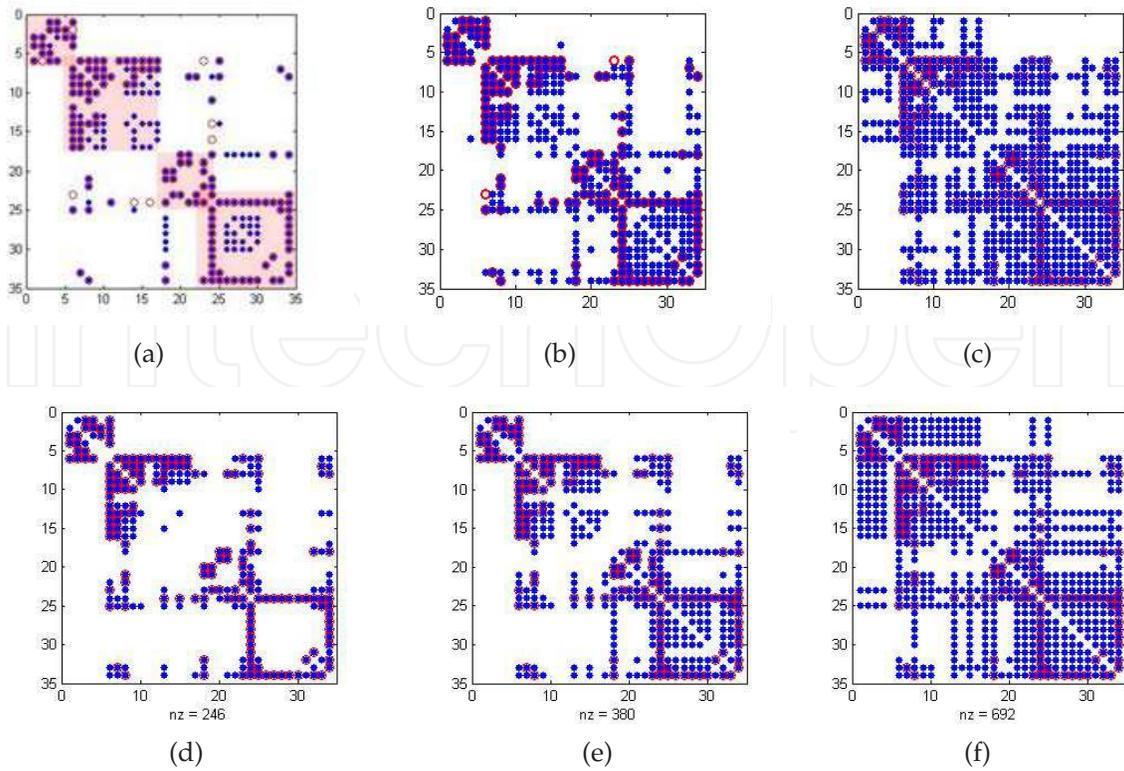


Fig. 9. Karate club: adjacency matrix is shown by red circles, detected communities by pink squares, predicted links are shown by blue dots. The upper part (a)-(c): predictions made by dynamical systems at different time scales. The bottom part (d)-(f): recommendations made by the modified Katz predictor at different values of β .

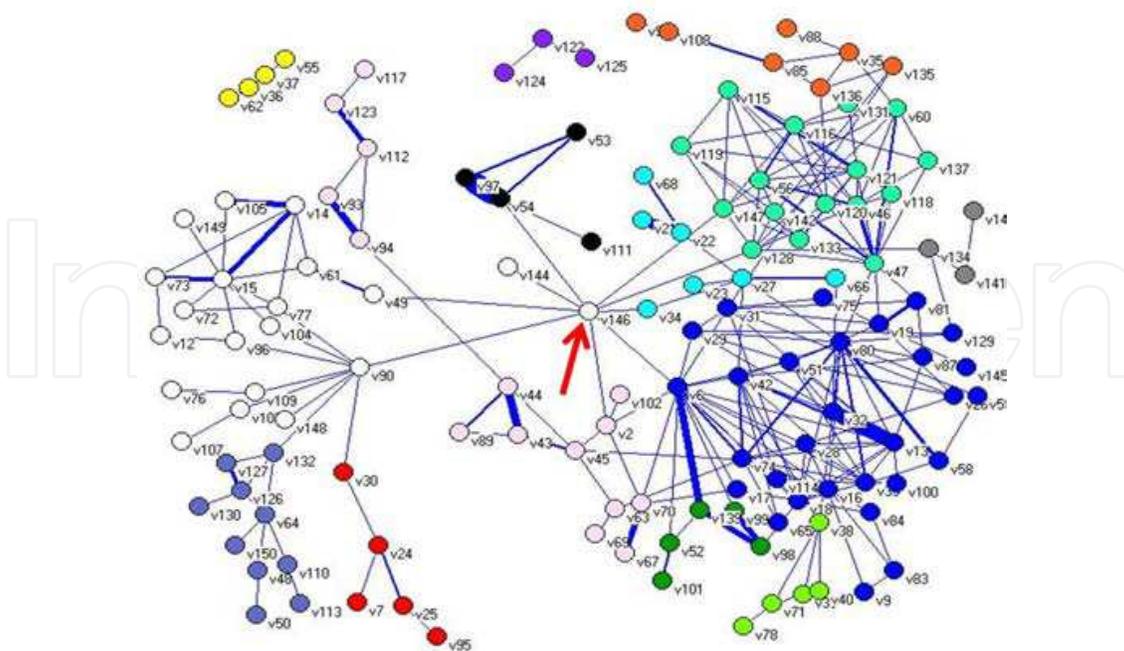


Fig. 10. Community detection based on SMS and call-logs: communities are coded by colors.

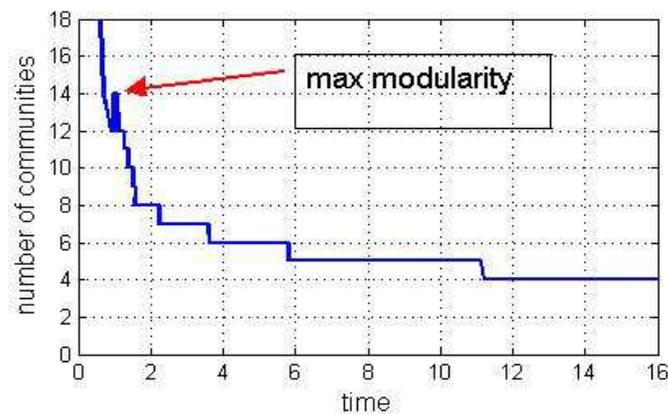


Fig. 11. Stability of communities at different resolution levels.

stable partitions appear at the higher hierarchical levels starting from 8 communities. In the following we rely on this fact to build the ground truth references for evaluation of clustering.

7.2 Applications for multi-layer graphs

Besides phone and SMS call-logs, the social affinity of participants may also be derived from other information layers, such as a local proximity of users (BT and WLAN layers) and their location information (GPS). In this case the soft communities detection may be extended to include multiple graph layers. In particular, we found that users' profiles may significantly vary across the layers. For example, a user may have dense BT connections with a multiple communities participation, while his phone call activities may be rather limited. Combining information from several graph layers can be used to improve the reliability of classification. Below we show some preliminary results, more detailed analysis of multi-layer graphs built from mobile datasets may be found in (Dong et al, 2011).

To make verification of detected communities we select a subset of 136 users with known email affiliations as the ground truth. In our case these users are allocated into 8 groups. To get the same number of communities in social affinity multi-layer graphs, we use the random walk (11) to obtain the more coarse resolution than provided by the modularity maximization. Fig 12 depicts communities (color coded) derived from the phone-calls graph. Single nodes here indicate users which did not make phone calls to other participants of the data collection campaign. Communities derived from the BT-proximity graph and mapped on the phone-call graph are shown at Fig.13. As expected, multi-layers graphs help us to classify users based on the additional information found in other layers. For example, users which can not be classified based on phone calls (Fig.12) are assigned to communities based on the BT proximity (Fig.13). Fig.14 shows communities detected in the combined graph formed by the BT and phone-call networks and then mapped on the phone-call network.

Next, we consider communities detected at single and combined layers with different strategies (35)-(37) described in Section 5 and compare them to the ground truth. To evaluate accuracy of community detection we use the normalized mutual information (NMI) score, purity test and Rand index (RI) (Manning et al, 2008). We found that the best graph combining is provided by weighted superposition (37) according to the *max*-modularity of layers Q . Results of the comparison are summarized in Table 1. As expected, different graph layers have a different relevance to the email affiliations and do not have fully overlapped community structures. In particular, the local proximity seems to be more relevant to professional relations

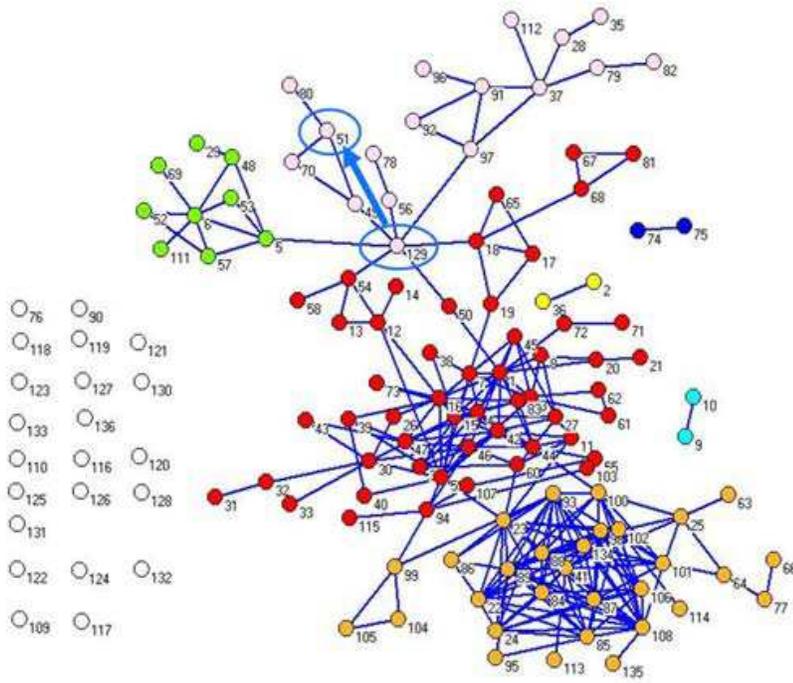


Fig. 12. Community detection using random walk in the phone-calls network.

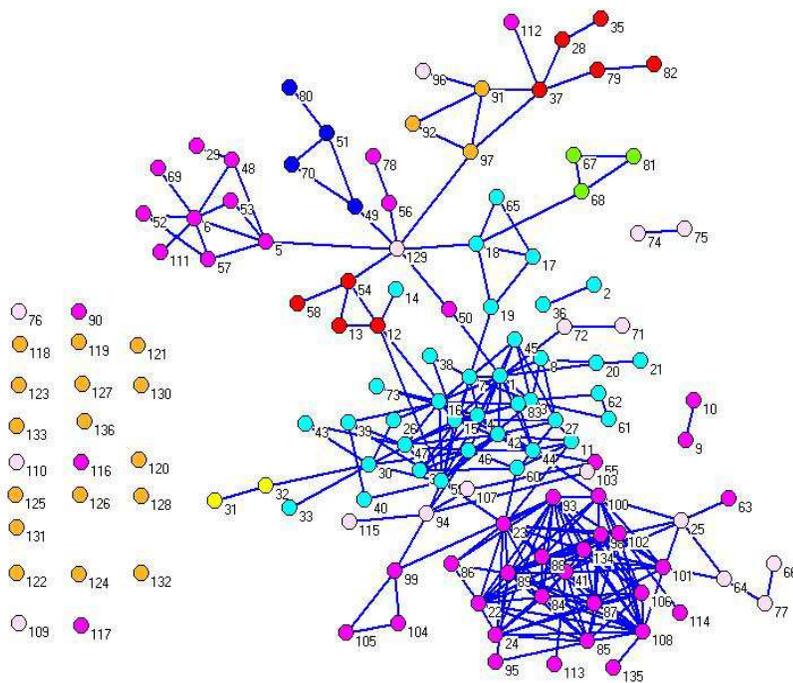


Fig. 13. Communities detected in the BT proximity network and mapped on the phone-calls network.

indicated by email affiliations, while phone calls seem to reflect more friendship and family relations. However, the detected structures are still rather close to each other (cf. columns in Table 1) reflecting underlying social affinity. As one can see, by properly combining information from different graph layers we can improve the reliability of communities detection.

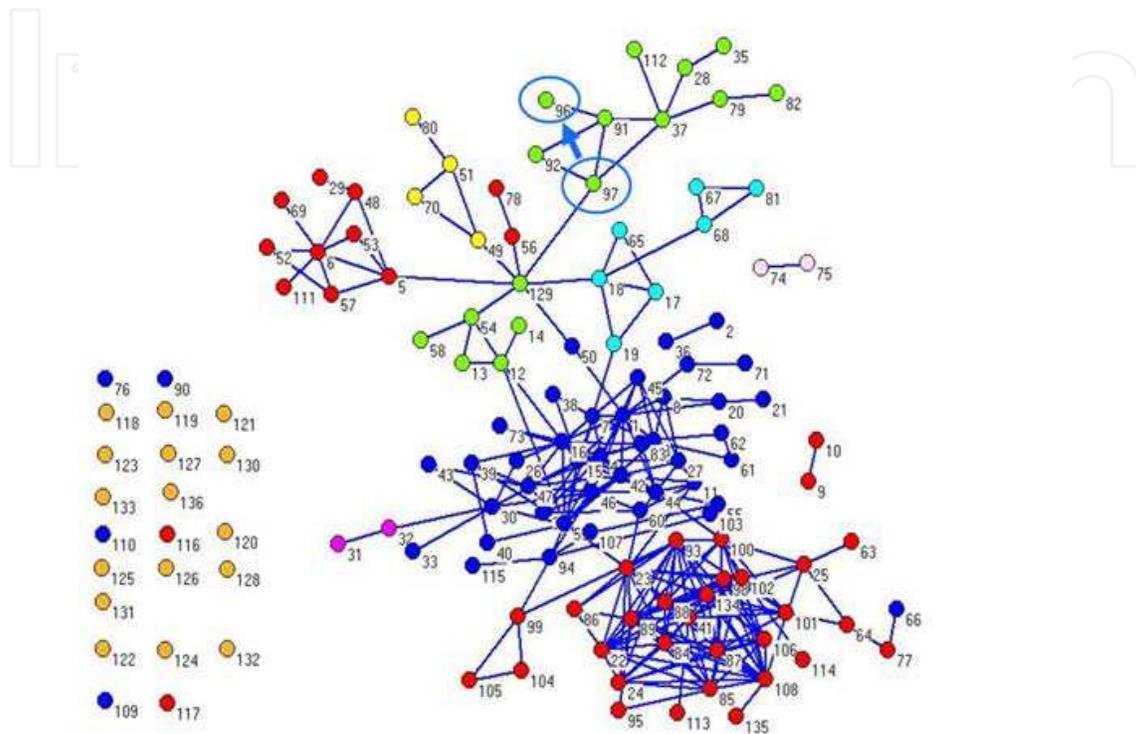


Fig. 14. Communities detected in the combined BT & phone-calls network and mapped on the phone-calls network.

	NMI	Purity	RI	Q
Phone calls	0.262	0.434	0.698	0.638
BT proximity	0.307	0.456	0.720	0.384
GPS	0.313	0.471	0.704	0.101
Phone + BT	0.342	0.427	0.783	

Table 1. Evaluation of community detection in multi-layer graphs.

7.3 Application for recommendation systems

As discussed in Section 4, one of applications of the soft communities detection and coupled systems dynamics may be seen in recommendation systems. To illustrate the approach we selected the user "129" (marked by oval) in the phone-calls network at Fig.12 and calculated proposed prediction scores for different similarity measures.

First, we consider intra-community predictions made by coupled dynamical systems. Fig.15(a) depicts pair-wise correlations (scaled by 5) between oscillators at $t = 10$ for the sub-network at Fig.12 forming the intra-community of the user "129". By changing

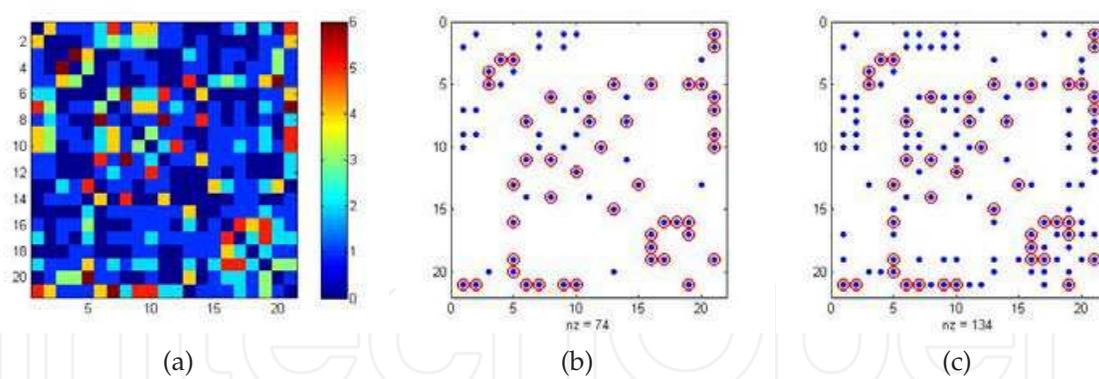


Fig. 15. Community of the user "129" (shown by pink color at Fig.12): averaged (scaled by 5) pair-wise correlations between oscillators at $t = 10$ (a). Intra-community adjacency matrix (red circles) and links predicted by dynamics (blue dots) at different resolution levels: $t = 15$ (b) and $t = 25$ (c).

the threshold η for the dynamical connectivity matrix $C_t(\eta)$ (which is linked to time resolution t) we obtain different connectivity matrices $C_\eta(t)$ presenting the network evolution. Connectivity matrices (blue points) corresponding to $\eta = 3$ ($t = 15$) and $\eta = 2.3$ ($t = 25$) are shown at Fig.15(b) and Fig.15(c), respectively. The community adjacency matrix is marked on the same figures by red circles. As one can see, dynamical systems first reliably detect the underlying topology and then form new links as the result of local interactions and dynamical links update. It can be easily verified that practically all new links (e.g., 12 out of 13 at Fig.15(b)) create new cliques, hence we can interpret these new links as the Friend-of-Friend recommendations.

Calculated scores $S_{DC}^{(i,i)}(k,n)$ for dynamical systems together with the Friend-of-Friend intra-community recommendations for two predictors based on the soft community detection (Katz predictor and convergence speed to consensus, $S_{CC}^{(i,i)}(k,n)$) are summarized in Table 2. Here we list all new links together with their normalized prediction scores for the user "129" which create at least one new clique within its community (shown by pink color at Fig.12).

source	destination	$S_{KC}^{(i,i)}(s,d), \%$	$S_{CC}^{(i,i)}(s,d), \%$	$S_{DC}^{(i,i)}(s,d), \%$
129	51	10.5	22.6	18.6
129	78	11.1	16.3	20.8
129	91	47.1	15.4	11.6
129	70	11.3	15.3	18.9
129	92	9.6	15.3	18.8
129	37	10.5	15.1	11.4

Table 2. Scores for the FoF intra-community recommendations for user 129 according to different similarity measures for the phone-calls network at Fig.12.

Recall that both $S_{CC}^{(i,i)}(k,n)$ and $S_{DC}^{(i,i)}(k,n)$ are based on the network synchronization with closely related Laplacians. As the result, the distribution of prediction scores $S_{CC}^{(i,i)}(k,n)$ and $S_{DC}^{(i,i)}(k,n)$ are rather close to each other, compared to the the distribution of the routing-based Katz score $S_{KC}^{(i,i)}(k,n)$. Convergence of opinions to a consensus within communities in many cases is the important target in social science. As an example, the best intra-community

recommendation in the phone-calls network according $S_{CC}^{(ii)}(k, n)$ is shown by the blue arrow at Fig.12. Scaled pair-wise correlations between oscillators for the whole phone-call network

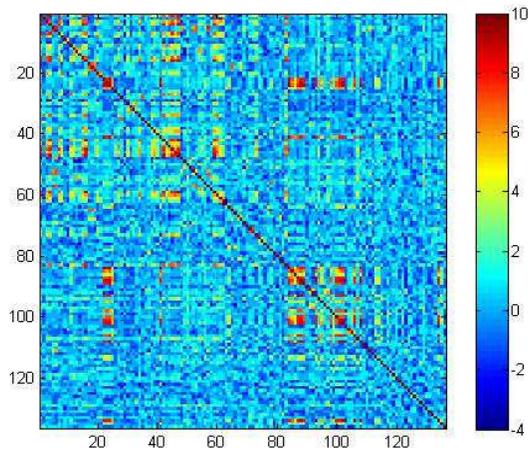


Fig. 16. Phone-call network: averaged pair-wise correlations (scaled by 10) between oscillators at $t=10$, coupling scenario $B.3$.

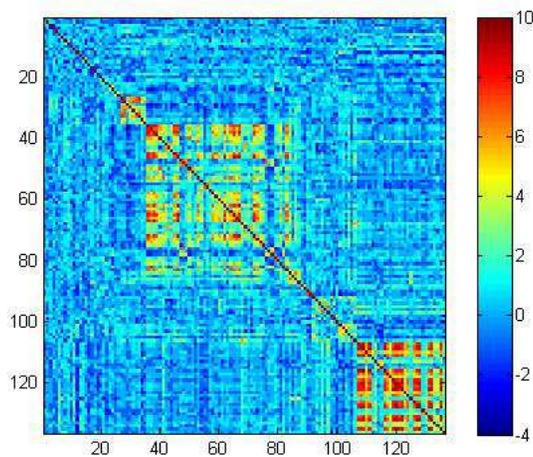


Fig. 17. Phone-call network: averaged pair-wise correlations re-ordered according to detected communities.

at Fig.12 are shown at Fig.16. Correlations between nodes, re-ordered according to one of the stable partitions detected by the random walk at $t=10$, reveal clearly the community structure (Fig.17). The phone-calls adjacency matrix (red circles) and all possible intra-community links (yellow squares) for the stable communities at $t = 10$ are depicted at Fig.18 (a). Links predicted by system dynamics (blue dots) inside and outside of yellow squares indicate predicted intra-community and inter-communities connections at different resolution levels and show the priority of the intra-community connections (Fig.18 (b) – Fig.18(c)). As the whole, the presented results for the coupled dynamical systems provide the formal basis for the recommendation rules formulated in Section 4.2.

As it is shown in Section 3, the dynamical process of opinions convergence may be seen as the first-order approximation of the network synchronization. At the same time, $S_{CC}^{(i,i)}(k, n)$

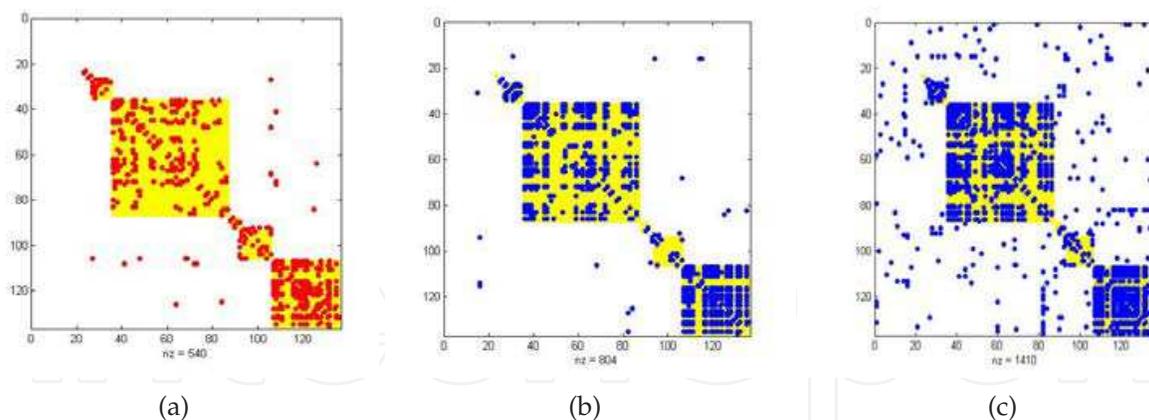


Fig. 18. Phone-call network: (a) adjacency matrix is marked by red dots, all possible intra-communities links are shown by yellow squares. Links predicted by dynamics (blue dots) tend to concentrate within communities: (b) $t = 10$; (c) $t = 15$.

has the lower computational complexity than $S_{DC}^{(i,i)}(k, n)$, it makes $S_{CC}^{(i,i)}(k, n)$ more suitable for large networks. Prediction scores $S_{CC}(129, n)$ and $S_{KC}(129, n)$ calculated according to (32) for cases with intra- and inter-communities links in the phone-call network are depicted at Fig.19. Here the scores are normalized as probabilities and sorted according to its priority; destination nodes n are listed along the x -axis; corresponding random-link probabilities, $p_{kn} = (d_k d_n) / 2m$, are shown as the reference. Note that the link with the highest priority, $\{129, 51\}$ for $S_{CC}^{(i,i)}(k, n)$, is the same as in the intra-community recommendation (cf. Table 2). However, the presence of inter-community links modifies priorities of other recommendations according to (30). To make verification we compare the predicted links at the phone-call network with links observed for the user "129" at the BT proximity layer. This comparison shows a good fit: 16 out of 18 predicted links are found at the BT proximity layer.

Results for the combined BT and phone-calls networks are presented at Fig.20. Pair-wise correlations between nodes obtained by dynamical systems approach are shown at Fig.20 (a). These correlations may be interpreted as probabilities for new links recommendations. Fig.20 (b) depicts recommended links based on the modified Katz predictor (blue circles) beyond the

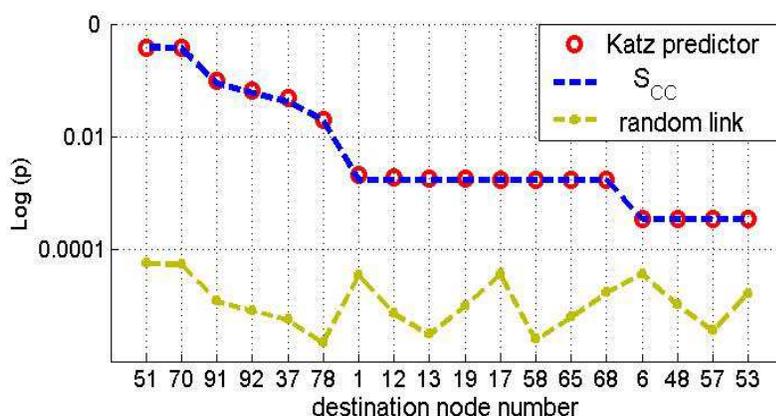


Fig. 19. Priorities of the FoF recommendations for the user 129 at Fig.12 to be connected to destination nodes shown along x -axis over all relevant communities.

given topology (red dots). We found that both recommenders mostly agree on the priority of intra-community links, but put different weights on inter-community predictions.

Depending on a purpose of recommendation we may select different prediction criteria. Since new links change topology, which in turn affects dynamical properties of the network, the recommendations may be seen as a distributed control driving the network evolution.

In general, the selection of topology-based recommendation criteria and their verifications are the open problems. Currently we are running experiments to evaluate different recommendation criteria and its acceptance rates.

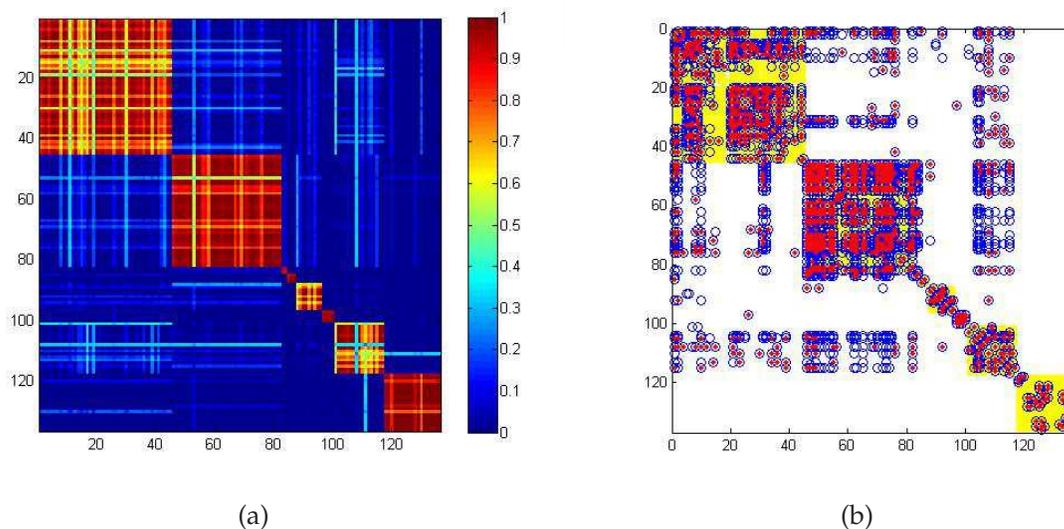


Fig. 20. Combined BT and phone-call networks, nodes are ordered according to detected communities: (a) color-coded pair-wise correlations using dynamical systems; (b) links recommendations using modified Katz predictor (blue circles), adjacency matrix is marked by red dots, all possible intra-community links are shown by yellow squares.

8. Conclusions

In this chapter we present the framework for multi-membership communities detection in dynamical multi-layer graphs and its applications for links predictions/recommendations based on the network topology. The method is based on the dynamical formulation of modularity using a random walk and then extended to coupled dynamical systems to detect communities at different hierarchical levels. We introduce attractive and repulsive coupling and dynamical link updates that allow us to make predictions on a cooperative or a competing behavior of users in the network and analyze connectivity dynamics.

To address overlapping communities we suggest the method of soft community detection. This method may be used to improve marketing efficiency by identifying users which are strongly relevant to targeted groups, but are not detected by the standard community detection methods. Based on the soft community detection we suggest friend-recommendations in social networks, where new link recommendations are made as intra- and inter-clique communities completion and recommendations are prioritized according to similarity measures modified to include multiple-communities membership. This developed methods are applied for analysis of datasets recorded during Nokia

mobile-data collection campaign to derive community structures in multi-layer graphs and to make new link recommendations.

9. Appendix: Clustering evaluation measures

Let's define $C = \{c_1, \dots, c_M\}$ and $\Psi = \{\psi_1, \dots, \psi_M\}$ as partitions containing detected clusters c_i and the ground truth clusters ψ_i , respectively. Quality of clustering algorithms may be evaluated by different measures (Manning et al, 2008), in particular:

- Rand index:

$$RI = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + FalseNegative + TrueNegative}; \quad (38)$$

- Purity test:

$$Purity(\Psi, C) = \frac{1}{n} \sum_{m=1}^M \max_j |\psi_m \cap c_j|; \quad (39)$$

- Normalized mutual information:

$$NMI(C, \Psi) = \frac{2 I(\Psi, C)}{H(\Psi) + H(C)}, \quad (40)$$

where the mutual information $I(C_1, C_2)$ between the partitions C_1 and C_2 and their entropies $H(C_i)$ are

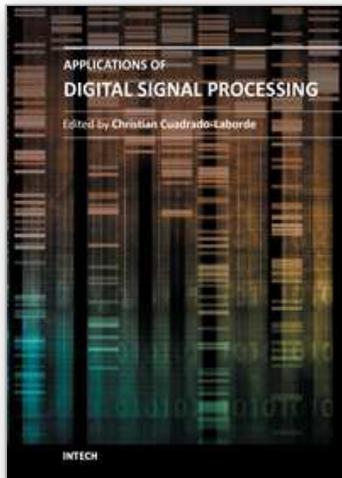
$$I(C_1, C_2) = \sum_{m_1}^M \sum_{m_2}^M \frac{c_{m_1, m_2}}{n} \log \left(\frac{n c_{m_1, m_2}}{n_{m_1} n_{m_2}} \right), \quad H(C_i) = - \sum_{m_i}^M \frac{n_{m_i}}{n} \log \left(\frac{n_{m_i}}{n} \right); \quad (41)$$

n is total number of data points; c_{m_1, m_2} is the number of common samples in the m_1 -th cluster from C_1 and the m_2 -th cluster in the partition C_2 ; n_{m_i} is the number of samples in the m_i -th cluster in the partition C_i . According to (41), $\max NMI(C_1, C_2) = 1$ if $C_1 = C_2$.

10. References

- Acebrón, J., Bonilla, L., Pérez-Vicente, C., Ritort, F., Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77 (1), pp. 137–185.
- Albert, R. & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, pp. 47–97.
- Arenas A., Díaz-Guilera, A., Pérez-Vicente, C. (2006). Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96, 114102.
- Arenas, A., Diaz-Guilera, A., Kurths, J., Moreno, Y. and Zhou, C. (2008). Synchronization in complex networks, *Physics Reports*, 469, pp. 93–153.
- Blondel, V., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, vol. 1742-5468, no. 10, pp. P10008+12.
- Evans, T. S. and Lambiotte R. (2009). Line Graphs, Link Partitions and Overlapping Communities. *Physical Review*, E 80 016105.

- Dong, X., Frossard, P., Vandergheynst, P. and Nefedov, N. (2011). Clustering with Multi-Layer Graphs: Spectral Perspective. *ArXiv*, 1106.2233.
- Flake, G., Lawrence, S., Giles, C. and Coetzee, F. (2002). Self-organization and identification of Web communities. *IEEE Computer* 35, pp. 66–71.
- Fortunato, S. (2011). Community detection in graphs. *Physics Reports*, 486, pp. 75–174.
- Girvan, M. & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99, pp. 7821–7826.
- Newman, M.E.J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review*, E 69, 026113.
- Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D. and Laurila, J. (2010). Towards Rich Mobile Phone Datasets: Lausanne Data Collection Campaign. *Proc. ACM Int. Conf. Pervasive Services*, Berlin.
- Kolda, T. and Bader, B. (2009). Tensor decompositions and applications, *SIAM Review*, vol.51, pp. 455–500.
- Kuramoto, Y. (1975). *Lectuer Notes in Physics*, 30, Springer NY.
- Lambiotte, R., Delvenne, J.-C. and Barahona, M. (2009). Laplacian Dynamics and Multiscale Modular Structure in Networks. *ArXiv:0812.1770v3*.
- Liben-Nowel, D. and Kleinberg, J. (2003). The Link Prediction Problem for Social Networks. *ACM Int. Conf. on Information and Knowledge Management*.
- Manning, C., Raghava, P. and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review*, E 69, 066133.
- Olfati-Saber, R. et al. (2007). Consensus and Cooperation in Networked Multi-Agent Systems. *IEEE Proceedings*, 95(1), pp. 215–233.
- Strehl A. & Ghosh, J. (2002). Cluster Ensembles - A Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of Machine Learning Research*, 3, pp. 583–617.
- Tang, L., Wang, W. and Wang X. (2009). Uncovering Groups via Heterogeneous Interaction Analysis. *SDM workshop on Analysis of Dynamic Networks*.
- Wasserman, S. & Faust, K. (1994). *Social Network Analysis*, Cambridge University Press, Cambridge.
- Zachary, W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33, pp. 452–473.



Applications of Digital Signal Processing

Edited by Dr. Christian Cuadrado-Laborde

ISBN 978-953-307-406-1

Hard cover, 400 pages

Publisher InTech

Published online 23, November, 2011

Published in print edition November, 2011

In this book the reader will find a collection of chapters authored/co-authored by a large number of experts around the world, covering the broad field of digital signal processing. This book intends to provide highlights of the current research in the digital signal processing area, showing the recent advances in this field. This work is mainly destined to researchers in the digital signal processing and related areas but it is also accessible to anyone with a scientific background desiring to have an up-to-date overview of this domain. Each chapter is self-contained and can be read independently of the others. These nineteenth chapters present methodological advances and recent applications of digital signal processing in various domains as communications, filtering, medicine, astronomy, and image processing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Nikolai Nefedov (2011). Multiple-Membership Communities Detection and Its Applications for Mobile Networks, Applications of Digital Signal Processing, Dr. Christian Cuadrado-Laborde (Ed.), ISBN: 978-953-307-406-1, InTech, Available from: <http://www.intechopen.com/books/applications-of-digital-signal-processing/multiple-membership-communities-detection-and-its-applications-for-mobile-networks>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen