Technical Report

UCAM-CL-TR-767
ISSN 1476-2986

Number 767

UNIVERSITY OF CAMBRIDGE

Computer Laboratory

Ising model of rumour spreading in interacting communities

Massimo Ostilli, Eiko Yoneki, Ian X. Y. Leung,
Jose F. F. Mendes, Pietro Lió, Jon Crowcroft

January 2010

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

http://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

http://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Ising Model of Rumour Spreading in Interacting Communities

Massimo Ostilli[a] [b], Eiko Yoneki[c], Ian X. Y. Leung[c], Jose F. F. Mendes[a], Pietro Lió[c], and Jon Crowcroft[c]

[a]Departamento de Física da Universidade de Aveiro, 3810-193 Aveiro, Portugal

[b]Statistical Mechanics and Complexity Center (SMC), INFM-CNR, Italy

[c]University of Cambridge Computer Laboratory, Cambridge, CB3 0FD, United Kingdom

We report a preliminary investigation on interactions between communities in a complex network using the Ising model to analyse the spread of information among real communities. The inner opinion of a given community is forced to change through the introduction of a unique external source and we analyse how the other communities react to this change. We model two conceptual external sources: namely, "Strong-belief", and "propaganda", by an infinitely strong inhomogeneous external field and a finite uniform external field, respectively. In the former case, the community changes independently from other communities while in the latter case according also to interactions with the other communities. We apply our model to synthetic networks as well as various real world data ranging from human physical contact networks to online social networks. The experimental results using real world data clearly demonstrate two distinct scenarios of phase transitions characterised by the presence of strong memory effects when the graph and coupling parameters are above a critical threshold.

1. Introduction

The prevalence of portable wireless technologies (e.g. social network services) has given rise to physical networks in the actual physical space along with online communications. The understanding of such pervasive network environment as one which is dynamic and pertaining to human subjects is an open research area The social relationships and interactions between human subjects bear significant impact to the design of such networks. Important results in areas such as complex network theory [1] has given various insights into this new generation of networks (i.e. social context).

A social network consists of a set of people forming socially "meaningful" relationships, such as those of acquaintance or physical co-location. In our context, we concern relationships where prominent patterns or information flow are observed. Studies has shown the various interesting statistical properties in social networks such as the well known small-world property and power-law degree distribution Notably, it is observed that human society (and in fact many other real world networks) naturally divides into groups or clusters, often referred to as communities. Community structure is an important attribute to the understanding of human social networks and plays a pivotal role in almost all networks across disciplines. Not only does it help us to understand the network itself, it is crucial in the design of strategies for information propagation and leads to realistic modelling of opinion dynamics in social networks. Community detection [20] aims to uncover efficiently the underlying community structure in a network.

In our previous work, we investigated the community structure in networks of human connectivity based on real world mobility traces [13,14,24] and in large-scale social networks [16]. Mostly, however, these approaches were empirical and heuristic-based focused on a single aggregated static network structure (i.e. community). For dynamic graph mining, Berger-Wolf et. al. show the study of community evolution based on node overlapping [2]. Further understanding of network model is essential in that important properties of human contact networks – such as community and weight of interactions – are important aspects of epidemics in information spread.

We attempt to address analytically the following questions in this paper. How does a community's opinion influence other communities? How long does it take to propagate information from one community to another? We explore an Ising model approach to understand the dynamics of information flow such as rumour spreading between communities in a network.

The last decade has seen numerous proposed sophisticated dynamics [4], however, the Ising model in all its variants has not been applied to social dynamics. To our knowledge, the Ising approach has remained unexplored in networks exhibiting community structure. Our general point of view is not far from that in [18]: in both approaches the external action of media or random events surrounding a community of persons is encoded in a suitable external field. Apart from the fact that in [18] the analysis is for one single community, the difference between the two models is the fact that in ours the disorder concerns the couplings and the links, whereas in [18] the source of the disorder lies in an additional external random field which mimics the randomness of the coupling between the external medias and the persons. [6] analysed the spread of a cultural treat: similar ideas were used for a case with $n = 2$, infinite connectivity, and no disorder.

In the following section, we propose the simplest version of the Ising model and discuss a simple example. We also show how to make the model more sophisticated towards more realistic situations in later sections. We demonstrate the analysed results with various real world data ranging from human physical contact networks to online social networks by applying the Ising interaction model.

2. The Model

Consider a network with n communities, we attempt to analyse how a rumour spreads out from, say, the community $\mathcal{C}^{(1)}$, to the other communities $\mathcal{C}^{(l)}$, $l = 2, \dots, n$ (upper indices refer to the community index). Any person (spin) can take just two values, "yes" or "no" (up or down) according to an Ising variable σ. At any instant, the percentage of "yes" in $\mathcal{C}^{(l)}$ is given by $(1 + s^{(l)})/2$, where $s^{(l)}$ is the meta-spin associated to $\mathcal{C}^{(l)}$

$$s^{(l)} \stackrel{def}{=} \frac{\sum_{i \in \mathcal{C}^{(l)}} \sigma_i}{N^{(l)}}, \quad l = 1, \dots, n, \tag{1}$$

where $N^{(l)}$ is the size of $\mathcal{C}^{(l)}$. Two important measures, especially for small samples, are the average over the realisations of the spins σ_i (thermal averages, $\langle \sigma_i \rangle$), and the average over the graph realisations (averaged over the disorder, $\overline{\sigma_i}$). These two combined averages will be indicated by $m^{(l)}$ and are the order parameters of the model:

$$m^{(l)} = \overline{\langle s^{(l)} \rangle}, \quad l = 1, \dots, n. \tag{2}$$

We indicate by $c^{(l,k)}$ the matrix of the connectivities. In general, given a randomly chosen vertex (person) i in $\mathcal{C}^{(l)}$, $c^{(l,k)}$ represents the average number of spins of $\mathcal{C}^{(k)}$ connected to i. Note that $c^{(l,k)}$ and $c^{(k,l)}$ are not independent, in fact, from their definitions, it is easy to see

that they must satisfy the following detailed balance

$$N^{(l)}c^{(l,k)} = c^{(k,l)}N^{(k)}. \tag{3}$$

We force the community $\mathcal{C}^{(1)}$ to change $m^{(1)}$ through either one of two possible kinds of sources acting solely on $\mathcal{C}^{(1)}$: an infinitely strong and inhomogeneous external field, or a uniform and finite external field. In the former case, the community $\mathcal{C}^{(1)}$ will change independently from the other communities, while in the latter case $\mathcal{C}^{(1)}$ will change also according to the interaction with the other communities. In both cases, physically, the source of this change can be parameterised as a suitable external field $h^{(1)}$ (non homogeneous for the former case) acting only on $\mathcal{C}^{(1)}$, however, we do not need to specify $h^{(1)}$, nor the original source generating $h^{(1)}$ which will evolve during some finite time interval, to which we are not interested either. Such sources can represent, *e.g.*, strong personal belief of the individuals of $\mathcal{C}^{(l)}$ in the former case, or the action of a media accessible only in $\mathcal{C}^{(1)}$ in the second case. We denote these two possible cases respectively by *Strong-belief* and *Propaganda*. Intermediate situations are of course also interesting but in this work we will consider only the two aforementioned cases on the grounds that: *i)* they are simpler than the general case, *ii)* the key properties of communication among the communities remain valid in these simpler cases.

The change of opinions in $\mathcal{C}^{(1)}$ affects the opinions in the other communities according to the Ising model, i.e., persons accommodate their opinions in such a way to maximise the number of weighted pairwise-agreements within (intra) each community and between (inter) each couple of communities according to minimum of the following Hamiltonian

$$H = -\sum_{l}\sum_{i,j\in\mathcal{C}^{(l)}} J^{(l,l)}_{(i,j)}\sigma_i\sigma_j - \sum_{l<k}\sum_{i\in\mathcal{C}^{(l)},j\in\mathcal{C}^{(k)}} J^{(l,k)}_{(i,j)}\sigma_i\sigma_j, \tag{4}$$

where $J^{(l,l)}_{(i,j)}$ and $J^{(l,k)}_{(i,j)}$, for $l,k = 1,\ldots,n$ with $l \neq k$, are some intra and inter couplings, respectively. More precisely, when the temperature T is finite, the ising configurations are distributed around the zero-equilibrium ones (*i.e.*, the configurations that give the minimum for H) according to the Gibbs-Boltzamann distribution $p \propto \exp(-H/T)$.

3. *Strong-belief* case

3.1. Example with 2 communities at zero temperature

Suppose now we have $n = 2$ communities with connectivities $c^{(l,k)}$, $l,k = 1,2$. All we want to measure is how $m^{(2)}$, seen as a function of $m^{(1)}$, changes:

$$m^{(2)} = m^{(2)}(\text{control parameters}; m^{(1)}). \tag{5}$$

In the simplest version of the model with only positive couplings and at zero temperature (meaning there are no thermal fluctuations), an effective field theory [21] provides us the following system of equations

$$\begin{cases} m^{(1)} = \tanh(c^{(1,1)}m^{(1)} + c^{(1,2)}m^{(2)} + \text{external field}), \\ m^{(2)} = \tanh(c^{(2,1)}m^{(1)} + c^{(2,2)}m^{(2)}). \end{cases} \tag{6}$$

As it is sufficient for us to know that there exists a suitable non-homogeneous external field whose variation produces a variation of $m^{(1)}$ in all the range $[-1,1]$, we are not interested in the specification the external field acting on $\mathcal{C}^{(1)}$. We take into account only the fact that - at zero temperature - any infinitesimal external field (uniform or not) acting on $\mathcal{C}^{(1)}$ constrains completely the spins of $\mathcal{C}^{(1)}$, therefore the first equation of the system (6) gives a value for $m^{(1)}$

which is completely independent of both $m^{(2)}$ and the connectivities. We are therefore left with a single equation for $m^{(2)}$ seen as a function of the independent variable $m^{(1)}$:

$$m^{(2)} = \tanh(c^{(2,1)}m^{(1)} + c^{(2,2)}m^{(2)}). \tag{7}$$

The general structure of Eq. (7), with $m^{(1)}$ seen as an independent variable, is related to the naive mean-field approximation See Section 6, also known as Curie-Weiss mean-field equation. In our case, if we interpret $m^{(1)}$ as an external field, then from Eq. (7) we see that the total external field acting on $\mathcal{C}^{(2)}$ is amplified by the factor $c^{(2,1)}$.

From the solution of Eq. (7), by varying the control parameters $c^{(2,1)}$ and $c^{(2,2)}$, we can find different possible scenarios. In particular, for $c^{(2,2)} < 1$, the community $\mathcal{C}^{(2)}$ needs a non-zero value of $c^{(2,1)}m^{(1)}$ in order to have a non-zero value of $m^{(2)}$, while for $c^{(2,2)} > 1$ the community $\mathcal{C}^{(2)}$ is able to cooperate internally for a non-zero value of $m^{(2)}$ even without any external solicitation coming from $\mathcal{C}^{(1)}$ ($c^{(2,1)}m^{(1)} = 0$). The critical parameter $c^{(2,2)} = 1$ coincides of course with the the well known percolation threshold result of the Erdös-Rényi random graph [9] valid for one single graph (or community), that soon we will see the generalisation to the case of n communities with $n > 2$.

We report some examples in Figs. 1-3 (below the percolation threshold), and in Figs. 4-6 (above the percolation threshold). As it is evident from these figures, while the cases below the percolation threshold give a regular function for $m^{(2)}(m^{(1)})$, we see that the cases above the percolation threshold give a function $m^{(2)}(m^{(1)})$ with a discontinuity located at $m^{(1)} = 0$ with a jump of the order $O(1)$ (a first-order phase transition). Furthermore, above the percolation threshold, we see that there is an interval of values of $m^{(1)}$ centred in $m^{(1)} = 0$ where there exist 2 solutions for $m^{(2)}$ with opposite signs. In the thermodynamic limit, only the solution having $\mathrm{Sign}(m^{(2)}) = \mathrm{Sign}(m^{(1)})$ is stable (or more properly "leading"), while the other is a metastable state that can survive only for a finite time interval. In terms of social dynamics, the presence of a metastable state means that, when the persons have enough internal connectivity ($c^{(2,2)} > 1$), there is an inertia in changing their opinions, but when the inertia effect ends, an abrupt jump toward the opposite opinion takes place.

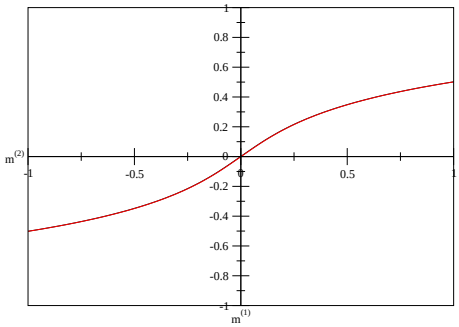

Figure 1. "Strong-belief" case. An example with $n = 2$, $c^{(2,1)} = 0.1$, $c^{(2,2)} = 0.9$.

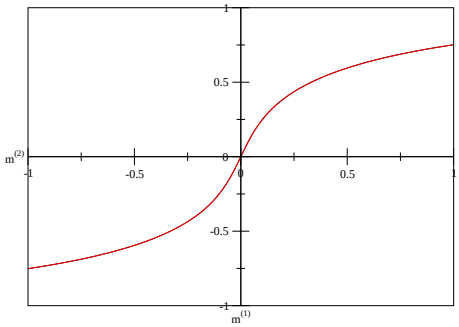

Figure 2. "Strong-belief" case. An example with $n = 2$, $c^{(2,1)} = 0.3$, $c^{(2,2)} = 0.9$.

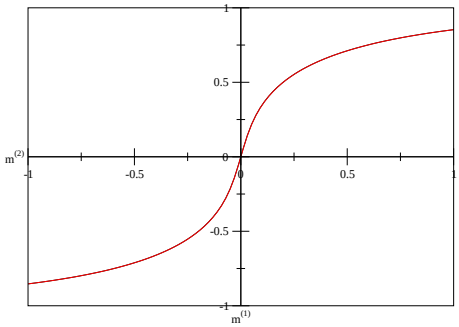

Figure 3. "Strong-belief" case. An example with $n = 2$, $c^{(2,1)} = 0.5$, $c^{(2,2)} = 0.9$.

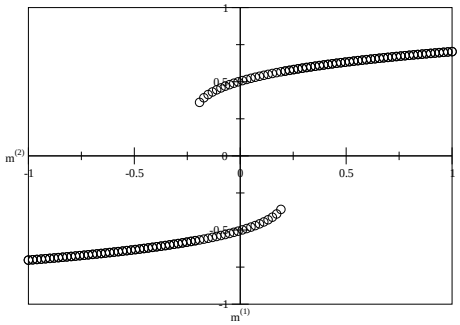

Figure 4. "Strong-belief" case. An example with $n = 2$, $c^{(2,1)} = 0.1$, $c^{(2,2)} = 1.1$.

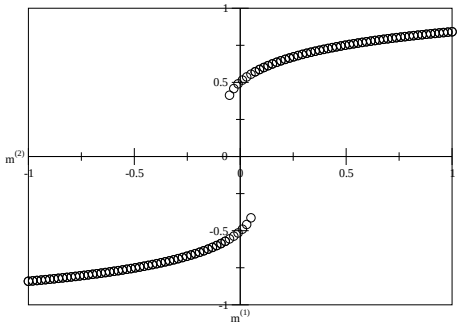

Figure 5. "Strong-belief" case. An example with $n = 2$, $c^{(2,1)} = 0.3$, $c^{(2,2)} = 1.1$.

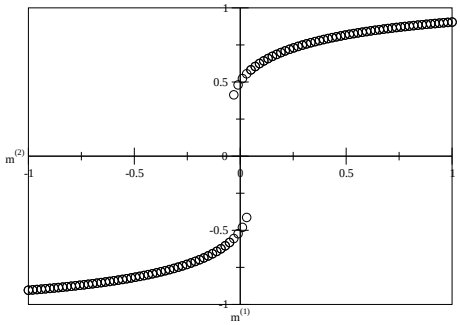

Figure 6. "Strong-belief" case. An example with $n = 2$, $c^{(2,1)} = 0.5$, $c^{(2,2)} = 1.1$.

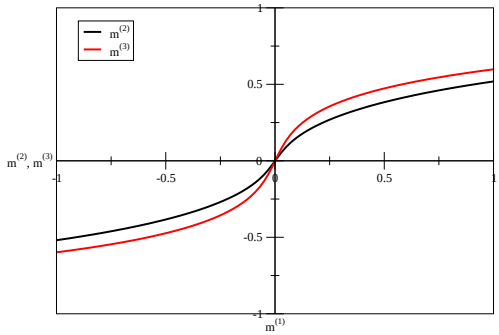

Figure 7. "Strong-belief" case. An example with $n = 3$, $c^{(2,1)} = c^{(3,1)} = c^{(2,3)} = c^{(3,2)} = 0.1$, $c^{(2,2)} = 0.8$, $c^{(3,3)} = 0.9$. In this case $\det(\mathbf{1} - \boldsymbol{c}) = 0.01 > 0$.

3.2. Generalisation to an arbitrary number of communities and finite temperature

The generalisation of the above scheme to the case with $n > 2$ communities is straightforward. For example, for $n = 3$, we have to solve the following system with respect to $m^{(2)}$ and $m^{(3)}$ both seen as functions of $m^{(1)}$:

$$\begin{cases} m^{(2)} = \tanh(c^{(2,1)}m^{(1)} + c^{(2,2)}m^{(2)} + c^{(2,3)}m^{(3)}) \\ m^{(3)} = \tanh(c^{(3,1)}m^{(1)} + c^{(3,2)}m^{(2)} + c^{(3,3)}m^{(2)}). \end{cases} \tag{8}$$

In general, given $n > 2$ and the reduced matrix of the connectivities $\boldsymbol{c}$ with dimension $(n-1) \times (n-1)$, it easy to see that the corresponding reduced system of $(n-1)$ equations in the unknowns $(n-1)$ order parameters $m^{(2)}, \ldots, m^{(n)}$, admits a non zero solution even for $m^{(1)} = 0$ as soon as the following condition is satisfied

$$\max\{\text{Eigenvalues of } \boldsymbol{c}\} > 1, \tag{9}$$

which in particular implies that the percolation critical surface is given by

$$\det(\mathbf{1} - \boldsymbol{c}) = 0. \tag{10}$$

Eq. (10) generalises the percolation threshold of the Erdös-Rényi random graph to the case of $n-1$ generic interacting graphs. When we are above such a critical surface means that, for example, one community is able to cooperate internally as in the previous case ($n = 2$) so that spins get aligned even without any external field. However, now ($n > 2$), it can also happen that two or more communities cooperate synergically by exploiting their inter-couplings; in other words, in the thermodynamic limit, a giant connected component can appear also as a result of a mutual cooperation among more communities. Mathematically speaking, this is a consequence of the fact that Eq. (10) is a single algebraic equation in $n-1$ unknowns (*i.e.*, the solution is a $n-2$ dimensional surface).

In Figs. 7 and 8 we report a case at zero temperature below and above the percolation threshold surface, respectively. As in the case $n = 2$ above the percolation surface there appear also metastable solutions. In general, the number of metastable solutions grows smoothly with n, however, if in the model some of the couplings are negative, the number of solutions may grow exponentially fast with n, as the system becomes a spin glass [17,11].

More in general the system can be analysed at finite temperatures. In this case, to keep $m^{(1)}$ rigidly independent from the other m's (the "Strong-belief" case), it is necessary that the external and inhomogeneous external field acting on $\mathcal{C}^{(1)}$ be infinite. Once again we stress however

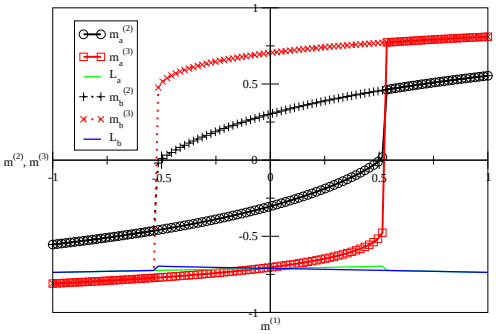

Figure 8. "Strong-belief" case. An example with $n = 3$, $c^{(2,1)} = c^{(3,1)} = c^{(2,3)} = c^{(3,2)} = 0.1$, $c^{(2,2)} = 0.8$, $c^{(3,3)} = 1.2$. In this case $\det(\mathbf{1} - \boldsymbol{c}) = -0.05 < 0$.

that we do not need to specify such a field. Let T be the temperature. Finite temperature implies that the spin configurations belonging to the other communities have some freedom to oscillate around the zero temperature equilibrium ones. In this case the system (8) modifies as

$$\begin{cases} m^{(2)} = \tanh\left(c^{(2,1)}t^{(2,1)}m^{(1)} + c^{(2,2)}t^{(2,2)}m^{(2)} + c^{(2,3)}t^{(2,3)}m^{(3)}\right) \\ m^{(3)} = \tanh\left(c^{(3,1)}t^{(3,1)}m^{(1)} + c^{(3,2)}t^{(3,2)}m^{(2)} + c^{(3,3)}t^{(3,3)}m^{(3)}\right), \end{cases} \tag{11}$$

where $\beta = 1/T$, and $\boldsymbol{t}$ is the matrix defined as

$$t^{(l,k)} \stackrel{def}{=} \tanh(\beta J^{(l,k)}), \tag{12}$$

where the matrix $J^{(l,k)}$, represents the matrix of the couplings. When we pass from zero to finite temperatures, the role of the percolation threshold passes to be played by the critical surface in the plane (T, c) whose equation becomes:

$$\det\left(\mathbf{1} - \boldsymbol{c_t}\right) = 0, \tag{13}$$

where the matrix $\boldsymbol{c_t}$ is given by

$$c_t^{(l,k)} = c^{(l,k)}t^{(l,k)}. \tag{14}$$

Examples with $n = 4$ for the local-strong case at finite temperature are reported in Figs. (20-23).

4. *Propaganda* case

In this case the community $\mathcal{C}^{(1)}$ is subjected to a finite and uniform external field h_1 so that it will be no completely constrained by the field and it will accommodate its configurations in order to be in equilibrium with all the other communities. Therefore, for the generic case of n communities with a generic temperature $T = 1/\beta$, and a couplings matrix $J^{(l,k)}$, for $l, k = 1, \ldots, n$, given a value of h_1, we will have to solve the following system with respect to all the n "order parameters" $m^{(1)}, m^{(2)}, \ldots, m^{(n)}$

$$\begin{cases} m^{(1)} = \tanh\left(c^{(1,1)}t^{(1,1)}m^{(1)} + \ldots + c^{(1,n)}t^{(1,n)}m^{(n)} + \beta h_1\right) \\ \ldots \\ m^{(n)} = \tanh\left(c^{(n,1)}t^{(n,1)}m^{(1)} + \ldots + c^{(n,n)}t^{(n,n)}m^{(n)}\right). \end{cases} \tag{15}$$

Then, if we want to analyse how $m^{(2)}, \ldots, m^{(n)}$ change as functions of $m^{(1)}$ we have to solve the system (15) for several values of h_1 and plot $m^{(2)}, \ldots, m^{(n)}$ versus $m^{(1)}$. The condition for the critical temperature is as Eq. (13). Note that unlike the "Strong-belief" case, the "Propaganda" case at zero temperature $\mathcal{C}^{(1)}$ will give $|m^{(1)}| \equiv 1$ for any $h_1 \neq 0$.

Examples with $n = 4$ of the "Propaganda" case at finite temperature are reported in Figs. (24-27).

Note that, from the physical point of view, in both the "Strong-belief" and "Propaganda" cases, $m^{(1)}$ plays exactly the same role of a term proportional to an external field acting on the other communities, the l-th one ($l \neq 1$) having a proportional factor given by $c_t^{(l,1)}$. However, while in the "Strong-belief" case $m^{(1)}$ takes any value in the range $[-1, 1]$, in the "Propaganda" case, near the critical temperature, $m^{(1)}$ cannot take all the values in the range $[-1, 1]$, since $m^{(1)}$ undergous finite jumps near the critical point. This is better seen in Fig. (28) where we plot the m's versus the external field h_1.

5. Real World Community Data

In this paper, we use three experimental datasets. One of them is gathered by the Haggle Project [12], referred to as "Cambridge"; another dataset from the MIT Reality Mining Project [8], referred to as "MIT", where proximity information with nearby nodes are discovered through periodic Bluetooth scans and location information from cell tower IDs. The last data is extracted from the social network of Facebook. We crawled two sub-networks from Stanford and Harvard Universities in the Facebook entries. For Harvard students, the first 35,000 Facebook profiles belong to past students who were the first users Approximately 20,000 users with 2.1 million links now exist. The mean degree is 63, and a maximum degree is 911. The users have 1,660 distinct affiliations, of which 1,181 could be mapped to geographic locations all over the globe. We extracted the entries, who has "Stanford University" as a secondary affiliation in the Harvard students as well as the entries, who has "Harvard University" as a secondary affiliation in the Stanford students. Thus, the first group indicates the people who moved from Harvard to Stanford while the second group indicates the people who moved from Stanford to Harvard.

Previously, the characteristics of these datasets such as inter-contact and contact distribution have been explored in several studies [5][25], to which we refer the reader for further background information. We believe these datasets cover a rich diversity of environments from a physical urban environment in university town (Cambridge) to an online social networks. See [3]for the detailed analysis of Harvard-Stanford data. Table 1 summarises the configuration.

Table 1. Characteristics of the experimental data

Experimental data set	MIT	CAM	Harvard-Stanford
Device	Phone	iMote	online
Network type	Bluetooth	Bluetooth	Facebook
Duration (days)	246	11	current
Granularity (seconds)	300	120	-
Number of Experimental Devices	97	36	663
Number of Communities	4	2	2
Average Size of Communities	7.5	16.5	331.0

MIT: in the MIT Reality Mining project [8], 97 smart phones were deployed to students and staff at MIT over a period of 9 months. These phones were running software that logged contacts.

CAM: in the Cambridge Haggle project [15], 36 iMotes were deployed to 1st year and 2nd year undergraduate students for 11 days. iMotes detect proximity using Bluetooth.

Harvard-Stanford: two sub-networks consisting of students from Stanford and Harvard universities are extracted from Facebook. From those sub-networks, only the entries, who has both affiliations, Harvard and Stanford are extracted. The number of nodes with Harvard and Stanford in order is 325 and the number of nodes with Stanford and Harvard in order is 337.

5.1. Community Detection Method

Many community detection methods have been proposed and examined in the literature (see the recent review papers by Newman [20] and Danon *et al.* [7]). We have shown various community detection mechanisms [13] which can be applied to human connectivity traces. In this paper, for the datasets of MIT and Cambridge, we have exploit both K-Clique and *Fiedler Clustering*, where *Fiedler Clustering* captures more hierarchical structure.

The connectivity traces can be represented by weighted graphs, in which the weight of an edge represents the *contact duration* and *contact frequency* for the two end vertices. Many real world networks are weighted. A weighted graph can be converted into a multi-graph with many unit edges. We built various weighted graphs before applying community clustering algorithms. We only consider symmetric edges, on the other hand, edges can be symmetric (undirected) or asymmetric (directed), possibly with a different strength in either direction. Harvard-Stanford dataset is simply divided into 2 clusters based on the order of the affiliation description of Harvard and Stanford.

- K-Clique: Palla *et al.*[23] define a community as a union of all k-cliques (complete subgraphs of size k) that can be reached from each other through a series of adjacent k-cliques, where two k-cliques are said to be adjacent if they share $k-1$ nodes. An advantage of this approach is that it allows overlapping communities, which is useful as, in human society, one person may belong to multiple communities.
- *Weighted Networks Analysis* by Newman [19] can work on weighted graphs built from encounter trace. A weighted graph can be converted into a multi-graph with many unit edges.
- *Fiedler Clustering*: The eigenvector for the nonzero smallest eigenvalue of a Laplacian matrix is called the Fiedler vector [10]. This vector can be used for decomposing graphs into structural components.

6. Applying Ising Model to Real Data with Two Communities

Let us consider for simplicity again the case with $n = 2$ community. If we have the connectivities matrix $\boldsymbol{c}$, (or in general the matrix $\boldsymbol{c_t}$ for finite temperature (14)) and the percentage of "yes" ($2 \times$ (% of yes) $-1 = m$) on the two communities taken at several distinct times, we can derive the histogram of $m^{(2)}$ versus $m^{(1)}$ and compare with our theoretical solution. It is important, however, for such a validation that the community $\mathcal{C}^{(2)}$ with respect to the feature "yes" or "not", apart from the connection with $\mathcal{C}^{(1)}$, to be completely isolated. For example, there is a global rumour ("Propaganda" or "Strong-belief") starting on $\mathcal{C}^{(1)}$ that Prof. Smith next year will win the Nobel Prize in Physics. Then, in order to find some validation for the case with

$n = 2$ communities, it is important that $\mathcal{C}^{(2)}$ can receive such a rumour only from $\mathcal{C}^{(1)}$: any other information travelling on different channels will invalidate the model with $n = 2$.

Of particular interest can be the comparison of the slope of $m^{(2)}$ with respect to $m^{(1)}$:

$$\frac{\partial m^{(2)}}{\partial m^{(1)}} = \frac{\tilde{\chi}^{(2,1)}}{\tilde{\chi}^{(1,1)}}, \tag{16}$$

where, in general, the matrix $\tilde{\boldsymbol{\chi}}$ is defined as

$$\tilde{\chi}^{(l,k)} \stackrel{def}{=} \frac{\partial m^{(l)}}{\partial \beta h^{(k)}}. \tag{17}$$

For $\tilde{\boldsymbol{\chi}}$ the theory predicts

$$\tilde{\boldsymbol{\chi}} = \tilde{\boldsymbol{\chi}}_{\mathbf{0}} \cdot \left(\mathbf{1} - \tilde{\boldsymbol{\chi}}_{\mathbf{0}} \cdot \boldsymbol{c_t}\right)^{-1}, \tag{18}$$

where

$$\tilde{\chi}_0^{(l,k)} = \left[1 - m^{(l)}\right]^2 \delta_{l,k}. \tag{19}$$

In particular, in the paramagnetic region ($m^{(l)} = 0$ for any l), Eqs. (18) and (19) give

$$\tilde{\boldsymbol{\chi}} = \left(\mathbf{1} - \boldsymbol{c_t}\right)^{-1}. \tag{20}$$

In Figs. 9 and 10 we report the model theoretical predictions at zero temperature for $n = 2$ where the connectivity matrix $\boldsymbol{c}$ comes from real data.

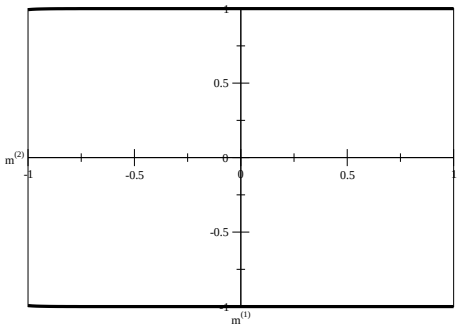

Figure 9. "Strong-belief" case at zero temperature. A case with $n = 2$ and matrix $c11 = 8, c12 = 10.78, c21 = 8.08, c22 = 11$. Cambridge students ($N = 36$) data-set analysed by using K-clique. Here we plot $m^{(2)}$ vs $m^{(1)}$.

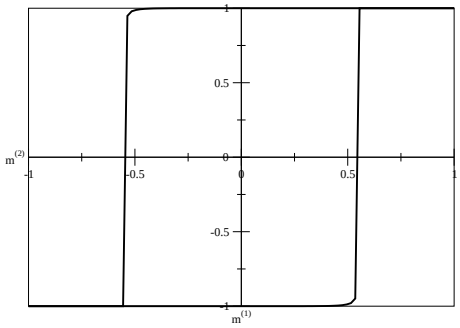

Figure 10. Data-set as Fig. 9. Here we plot $m^{(1)}$ vs $m^{(2)}$.

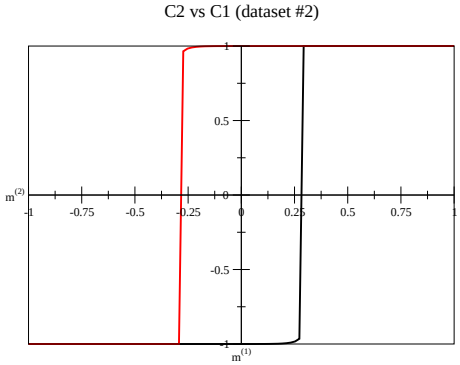

Figure 11. "Strong-belief" case at zero temperature. A case with $n = 4$ and matrix $c11 = 12, 12 = 6.92, c21 = 12.86, c22 = 5.72, c11 = 12, c13 = 3.77, c31 = 9.8, c33 = 4, c11 = 12, c14 = 3.47, c41 = 9.0, c44 = 4, c22 = 5.72, c23 = 2.29, c32 = 3.2, c33 = 4, c22 = 5.72, c24 = 1.0, c42 = 1.4, c44 = 4, c33 = 4, c34 = 5, c43 = 5, c44 = 4$. MIT students ($N = 97$) dataset analysed by using K-clique. Here we analyse the reduced system with $n = 2$ for $m^{(2)}$ vs $m^{(1)}$. For the case $m^{(1)}$ vs $m^{(2)}$ the plot is like Fig. 9.

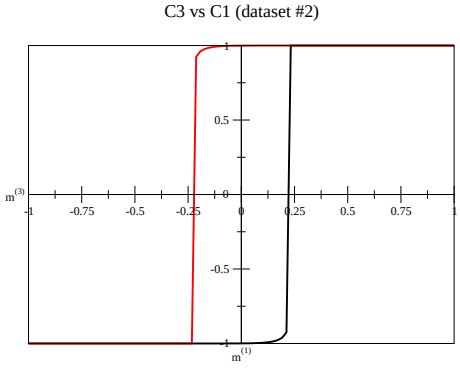

Figure 12. Data-set as Fig. 11. Here we analyse the reduced system $n = 2$ for $m^{(3)}$ vs $m^{(1)}$. Fo the case $m^{(3)}$ vs $m^{(1)}$ the plot is like Fig. 9.

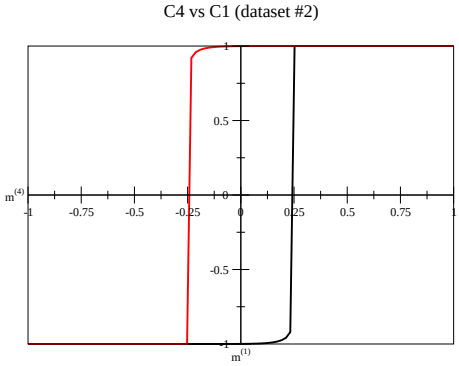

Figure 13. Data-set as Fig. 11. Here we analyse the reduced system $n = 2$ for $m^{(4)}$ vs $m^{(1)}$. Fo the case $m^{(3)}$ vs $m^{(1)}$ the plot is like Fig. 9.

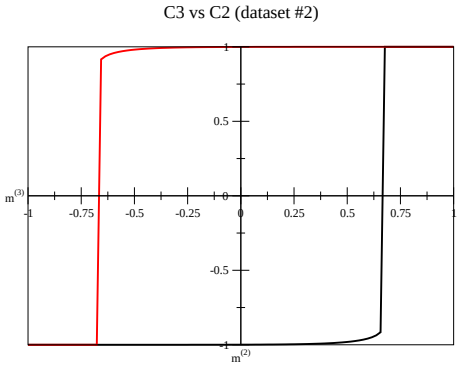

Figure 14. Data-set as Fig. 11. Here we analyse the reduced system $n = 2$ for $m^{(3)}$ vs $m^{(2)}$. Fo the case $m^{(2)}$ vs $m^{(3)}$ the plot is like Fig. 9.

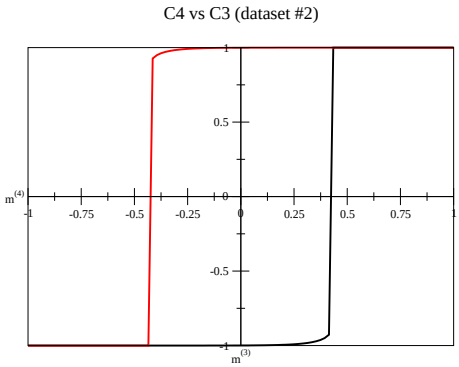

Figure 15. Data-set as Fig. 11. Here we analyse the reduced system $n = 2$ for $m^{(4)}$ vs $m^{(3)}$. Fo the case $m^{(3)}$ vs $m^{(4)}$ the plot is like Fig. 9.

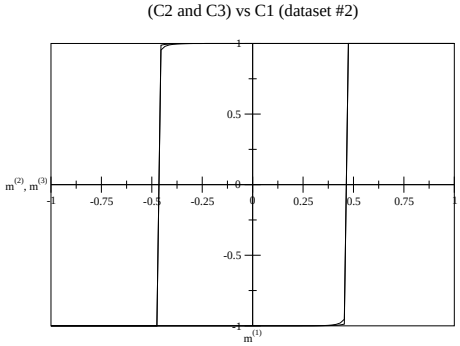

Figure 16. Data-set as Fig. 11. Here we analyse the reduced system $n = 3$ for $m^{(2)}$ and $m^{(3)}$ vs $m^{(1)}$. Fo the case $m^{(3)}$ vs $m^{(4)}$ the plot is like Fig. 9. In this case $m^{(3)}$ and $m^{(2)}$ coincide.

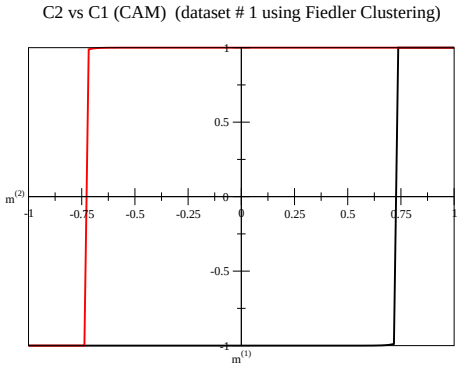

Figure 17. "Strong-belief" case at zero temperature. A case with $n = 2$. The dataset is the same as those of Figs. 9 and 10, but the matrix $\boldsymbol{c}$ is obtained by using the Fiedler clustering, being now $c^{(11)} = 18.8$, $c^{(12)} = 11.15$, $c^{(21)} = 14.87$, and $c^{(22)} = 13.4$. This plot must be compared with Fig. 9. The case $C1$ versus $C2$ produces instead a plot like Fig. 9, to be compared with Fig. 10. We see that the two different ways to obtain the matrix $\boldsymbol{c}$ produces here a sort of opposite situations; being, roughly speaking, the role of $C2$ and $C1$ reversed.

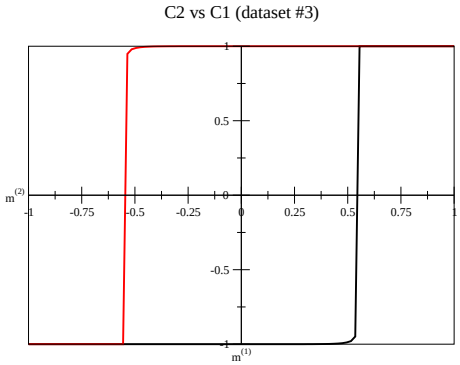

Figure 18. "Strong-belief" case at zero temperature. A case with $n = 2$, $c^{(11)} = 5.9$, $c^{(12)} = 5.9$, $c^{(21)} = 5.71$, and $c^{(22)} = 5.72$. Data-set taken from a Stanford-Harvard online social network ($N = 326 + 337$). For the case $C1$ versus $C2$ the plot is like Fig. 9.

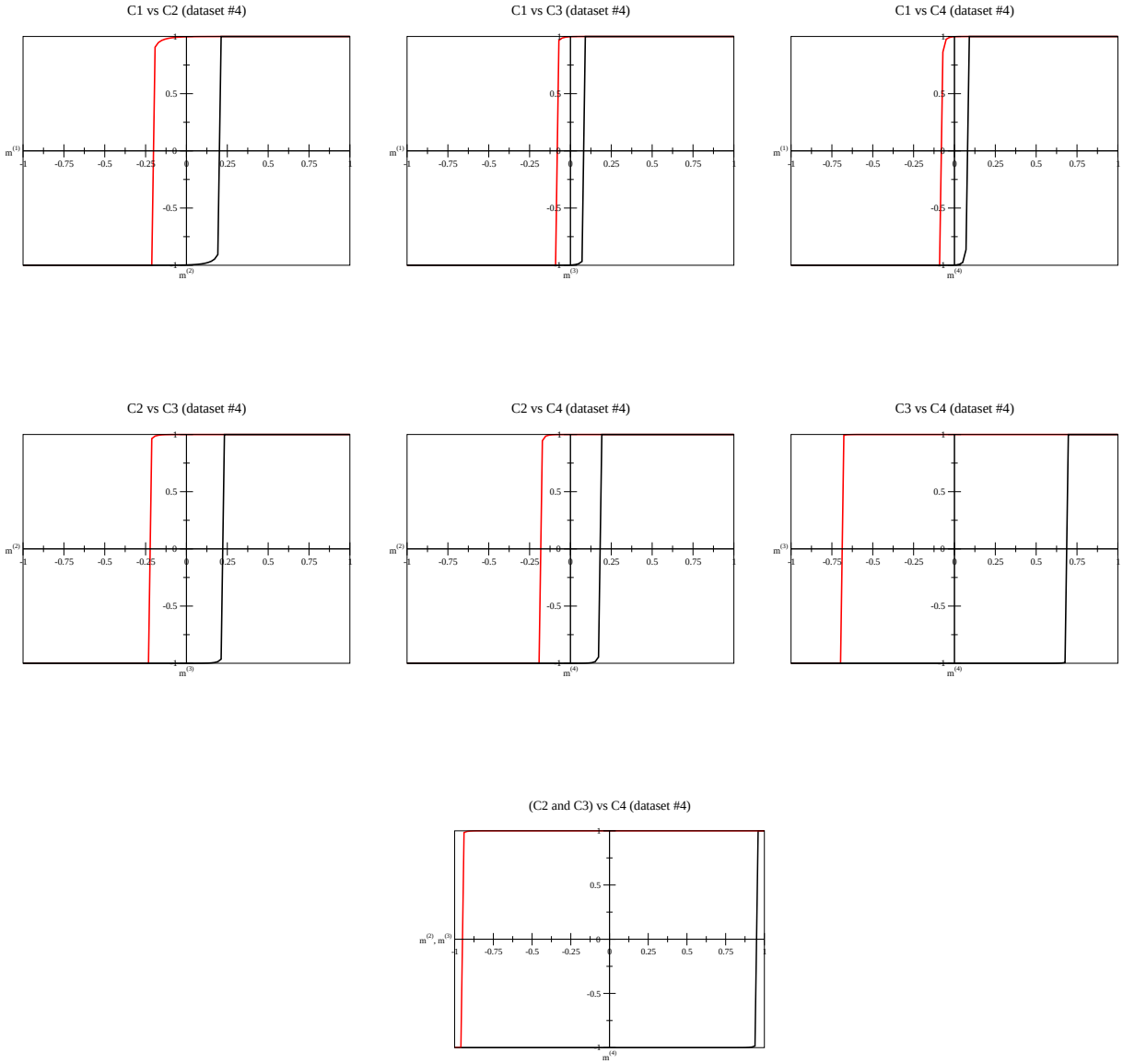

Figure 19. "Strong-belief" cases at zero temperature. The above seven plots are taken from a dataset having the following matrix: c11=3.32, c22=5.86, c33=18.72, c44=21.10, c12=7.83, c21=3.36, c13=19.3, c31=3.52, c14=24.5, c41=3.5, c23=17.21, c32=7.30, c24=21.86, c42=7.29, c34=23.18, c43=18.21. The data-set comes from the same set as Fig. 11 but it is analysed according to the Fiedler clustering algorithm. The cases not shown (including the cases with two or three interacting communities) are trivial, being these like Fig. 9. Such trivial cases are due to the fact that the non diagonal matrix elements are too small compared to very high values of the diagonal terms so that the system remains insensitive even to large variations of the independent variable $m^{(l)}$, where $l = 1, \ldots, 4$. Note that in the last plot (C^2 and C^3) vs C^4, the solutions $m^{(2)}$ and $m^{(3)}$ coincide. As is evident from the comparison with the above plots C^2 vs C^4, and C^3 vs C^4, from the plot (C^2 and C^3) vs C^4 we see that the communities C^2 and C^3 cooperate since their global inertia become bigger when their reciprocal inter-links are activated. The same argument applies to the case of Fig. 16.

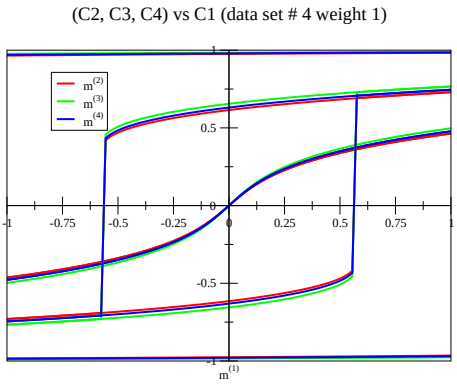

Figure 20. "Strong-belief" case. Analysis for a non-weighted network (w1) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to an external field which completely constrains $C^{(1)}$. The three groups of plots correspond to three different finite temperatures $T = 50$, $T = 40$, and $T = 20$ with coupling $J = 1$ (*i.e.* we are considering three situations with the adimensional couplings $\beta J = 0.02$, $\beta J = 0.025$, and $\beta J = 0.05$, respectively). Note that smaller temperatures correspond to steeper dependencies.

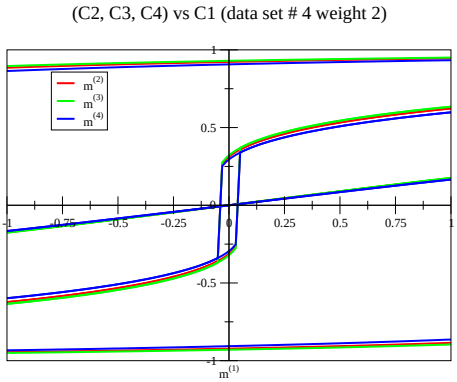

Figure 21. "Strong-belief" case. Analysis for a weighted network (w2) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to an external field which completely constrains $C^{(1)}$. The three groups of plots correspond to three different finite temperatures $T = 50$, $T = 33.3$, and $T = 20$ with coupling $J = 1$ (*i.e.* we are considering three situations with the adimensional couplings $\beta J = 0.02$, $\beta J = 0.03$, and $\beta J = 0.05$, respectively). Note that smaller temperatures correspond to steeper dependencies.

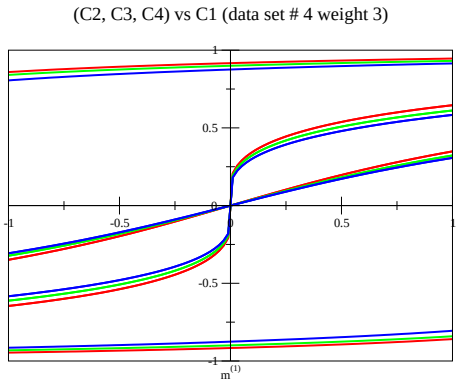

Figure 22. "Strong-belief" case. Analysis for a weighted network (w3) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to an external field which completely constrains $C^{(1)}$. The three groups of plots correspond to three different finite temperatures $T = 50$, $T = 20$, and $T = 12.5$ with coupling $J = 1$ (*i.e.* we are considering three situations with the adimensional couplings $\beta J = 0.02$, $\beta J = 0.04$, and $\beta J = 0.08$, respectively). Note that smaller temperatures correspond to steeper dependencies.

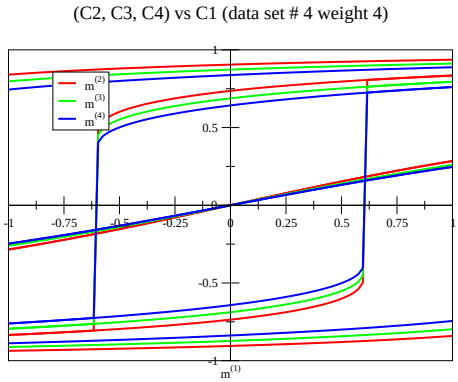

Figure 23. "Strong-belief" case. Analysis for a weighted network (w4) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to an external field which completely constrains $C^{(1)}$. The three groups of plots correspond to three different finite temperatures $T = 20$, $T = 12.5$, and $T = 10$ with coupling $J = 1$ (*i.e.* we are considering three situations with the adimensional couplings $\beta J = 0.05$, $\beta J = 0.08$, and $\beta J = 0.1$, respectively). Note that smaller temperatures correspond to steeper dependencies.

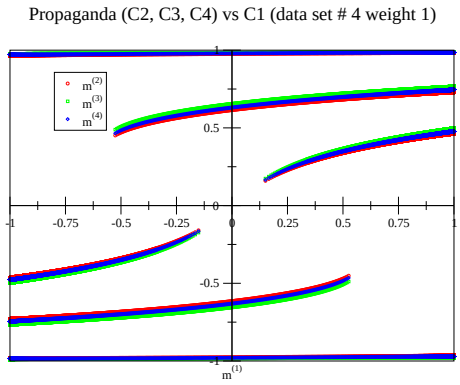

Figure 24. "Propaganda" case. Analysis for a weighted network (w1) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to a uniform external-propaganda field. The three groups of plots correspond to the same three different finite temperatures as in Fig. 20.

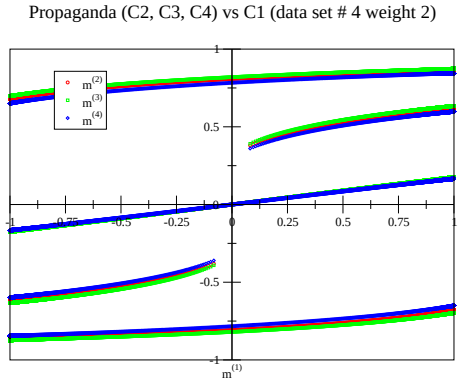

Figure 25. "Propaganda" case. Analysis for a weighted network (w2) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to a uniform external-propaganda field. The three groups of plots correspond to the same three different finite temperatures as in Fig. 21.

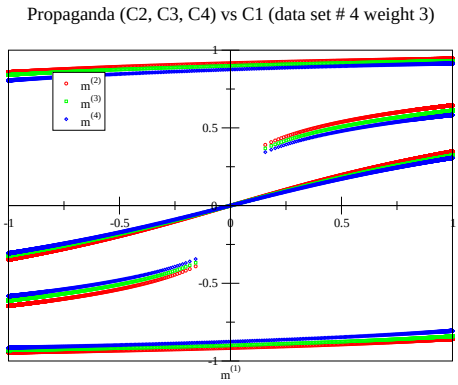

Figure 26. "Propaganda" case. Analysis for a weighted network (w3) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to a uniform external-propaganda field. The three groups of plots correspond to the same three different finite temperatures as in Fig. 22.

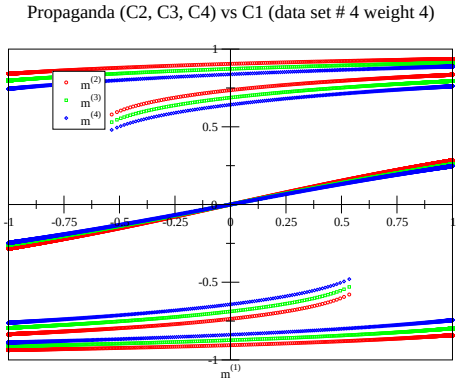

Figure 27. "Propaganda" case. Analysis for a weighted network (w4) at finite temperature. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of $m^{(1)}$, where $C^{(1)}$ (and only $C^{(1)}$) is subjected to a uniform external-propaganda field. The three groups of plots correspond to the same three different finite temperatures as in Fig. 23.

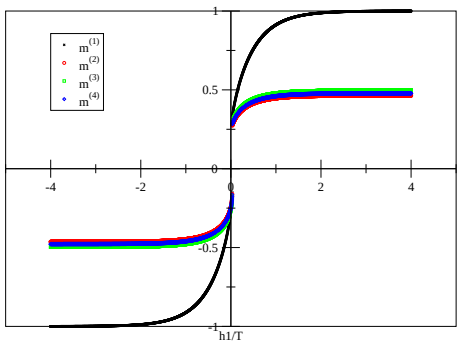

Figure 28. "Propaganda" case. Analysis for a weighted network (w1) at the finite temperature $T = 50$. The data-set comes from the same set as Fig. 11 and is analysed according to the Fiedler clustering algorithm. We plot $m^{(1)}$, $m^{(2)}$, $m^{(3)}$, $m^{(4)}$ as functions of βh_1, the external and uniform (propaganda) field acting only on $C^{(1)}$ (and only $C^{(1)}$).

The Cambridge dataset consists of two groups of under graduate students: one is the first year and the other is second year, where it demonstrate "Strong-belief" between the groups. Among MIT experiment participants, around 75% of participants are either student or faculty in the MIT Media Laboratory, while the remaining 25% is incoming students at the MIT Sloan business school next to the laboratory. Among the participants at the Media Lab, twenty are incoming masters students and five are incoming MIT new students. Thus, there could be 4 communities: 1) Media lab faculties, 2) Media lab students, 3) Business school faculties, and 4) Business school students. Thus, extracted four communities may be well correlated to these groups of participants. Nevertheless, cooperation between the community 2 and 3 is shown in 19 and two cases (i.e. "Strong-belief" and "Propaganda") are shown in the interaction between the community 2,3,4 against the community 3. Recently the reality mining project disclosed further collected information from the experiments and we will be able to explore more detailed social contexts in the MIT dataset for verification of our Ising modelling in future.

7. Conclusions and Implications

In this paper, we have reported preliminary investigations of interactions between communities using Ising spin model. Our modelling forces a given community to change its inner opinion through an external source acting only on it, and then we analyse how the other communities react to this change. There are two possible kinds of external sources: an infinitely strong and inhomogeneous external field, or a uniform and finite external field. In the former case (i.e. "Strong-belief"), the community will change independently from the other communities, while in the latter case the community will change also according to the interactions it has with the other communities (i.e. "Propaganda"). We have analysed various real world data ranging from human physical contact networks to online social networks by applying the Ising interaction model.

The experimental results using real world data clearly demonstrate two distinct scenarios of phase transitions. If we compare examples in Figs. 20 and 24, we see that the plots are almost identical but not near the critical T_c. In fact, near T_c, while in the "Strong-belief" case m_1 changes continuously, in the "Propaganda" case m_1 makes a finite jump from -0.25 to 0.25 and - correspondingly - m_2, m_3, and m_4 are forced to make a finite jump too. This is better shown in the example of Fig. 28 where we plot the m's as a function of the propaganda field h_1.

In the examples described in Section 2-6, the parameters of the model are encoded in the

single matrix $\boldsymbol{c}$ of the connectivities or, more generally, in the matrix $\boldsymbol{c_t}$ for finite temperatures. However, the version of the model we have so far considered is the simplest one: with only uniform and positive couplings, and no internal geometry. It is possible to consider generalisations to the model measuring the amplitude and the sign of the adimensional couplings, and to define two family of (adimensional) couplings, short-range, βJ_0, and long-range, βJ. The former are in correspondence with a possible geometry internal (to simplify) to the community, while the latter are in correspondence with random short-cuts of the underlying random graph. To realize these aim, a direct measure of the correlation function from the data should be done. Once we have access to the intra-correlation function (*i.e.*, inside the communities) we can evaluate βJ_0 and the diagonal matrix elements of the matrix $\boldsymbol{c_t}$, while for the inter-correlation function (*i.e.*, between two different communities) we get the non-diagonal matrix elements of $\boldsymbol{c_t}$.

When such generalisations are taken into account, a much more complex scenario of solutions and phase transitions are expected, as discussed in [21]. In particular, as already mentioned in the previous Section, when some of the couplings are negative, frustration takes place, and when the number n of interacting communities is large, a remarkable number of metastable states with a spin glass scenario will take place. Such situations are in fact expected in a social system, where an intrinsic unpredictability is a quite common situation.

Note that in Section 2, we have mentioned that the general structure of Eq. (7) is that of the Curie-Weiss mean field theory, which remains a naive approximation. However, the nature and the values of the "effective couplings" $c^{(2,1)}$ and $c^{(2,2)}$ is completely different and comes from the fact that the theory beyond Eqs. (6) is an effective mean-field theory that, besides the presence of possible short-range couplings, takes into account the fact that the kind of the disorder we take into account is quenched and not annealed. The method presented in [21] is limited being the proposed solution exact at any temperature only in the limits $\boldsymbol{c} \to 0$ or $\boldsymbol{c} \to \infty$. Nevertheless, it remains the only method to face such complicated models for $n \geq 2$. In particular, the method is always exact in the paramagnetic region and Eqs. (10) and (20) exact. See [22] for further details.

A direction for future research is to validate and verify the correctness of our modelling using experimental data from real world. We plan to use messages and advertisements for validation, which are often spread through online social networks. We find that choosing the parameter values of the network such as memory effects, which can be calculated analytically may become influential and we plan to verify various setting of parameters by real world experiments.

We also note here the importance of using weighted networks as a tool to differentiate the roles played by the communities. In fact, as it is evident from Figs. 20-27, by using increasing weights ($w1 \to w4$) we get larger and larger differences in the susceptibilities associated to different communities. Our current model is simple and we aim at extending this work to deal with more complex and realistic situations.

8. Acknowledgments

This research is funded in part by the EU Haggle project, IST-4-027918, and the SOCIALNETS project, 217141.

REFERENCES

1. R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47, 2002.
2. T. Berger-Wolf and J. Saia. A framework for analysis of dynamic social network. In *Proc. KDD*, 2006.

3. J. Bonneau, J. Anderson, F. Stajano, and R. Anderson. Eight friends are enough: Social graph approximation via public listings. In *Proc. ACM Workshop on Social Network Systems*, 2009.
4. C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.
5. A. Chaintreau et al. Impact of human mobility on the design of opportunistic forwarding algorithms. In *Proc. INFOCOM*, April 2006.
6. P. Contucci, I. Gallo, and S. Ghirlanda. *arXiv:0712.1119*.
7. L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas. Comparing community structure identification, 2005.
8. N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, V10(4):255–268, May 2006.
9. P. Erdös and A. Rényi. Statistical physics of social dynamics. *Publ. Math. Debrecen*, 6(290), 1959.
10. M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czech Math J*, 25, 1975.
11. K. Fischer and J. Hertz. *Spin Glasses*. Cambridge University Press, 1991.
12. Haggle Project. http://www.haggleproject.org, 2009.
13. P. Hui, J. Crowcroft, and E. Yoneki. BUBBLE Rap: Social Based Forwarding in Delay Tolerant Networks. In *MobiHoc*, 2008.
14. P. Hui, E.Yoneki, S. Chan, and J. Crowcroft. Distributed community detection in delay tolerant networks. In *Proc. MobiArch*, 2007.
15. J. Leguay et al. Opportunistic content distribution in an urban setting. In *ACM CHANTS*, 2006.
16. I. Leung, P. Hui, P. Lio', and J. Crowcroft. Towards real time community detection in large networks. *Phys. Rev. E*, 79(6), 2009.
17. M. Mezard, G. Parisi, and M. Virasoro. Spin glass theory and beyond. *Singapore: World Scientific*, 1987.
18. Q. Michard and J.-P. Bouchaud. *Eur. Phys. J. B*, 47(151), 2005.
19. M. Newman. Analysis of weighted networks. *Physical Review E*, 70:056131, 2004.
20. M. Newman. Detecting community structure in networks. *Eur. Phys. J. B*, 38:321–330, 2004.
21. M. Ostilli and J. F. F. Mendes. *J. Stat. Mech.*, 2009.
22. M. Ostilli and J. F. F. Mendes. Statistical physics of social dynamics. *Phys. Rev. E*, 80(011142), 2009.
23. G. Palla et al. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
24. E. Yoneki. Visualizing Communities and Centralities from Encounter Traces. In *ACM MobiCom - CHANTS*, 2008.
25. E. Yoneki, P. Hui, and J. Crowcroft. Wireless epidemic spread in dynamic human networks. *Bio-Inspired Computing and Communication*, LNCS(5151), 2008.