Deconstructing CSSS 2009 Social Network

Margreth Keiler
Department of Geography and Regional Research
University Vienna
Vienna, Austria
margreth.keiler@univie.ac.at

Murad Mithani
Rensselaer Polytechnic Institute
Troy, NY, USA
mithani.murad@gmail.com

Roozbeh Daneshvar
Department of Electrical Engineering
Texas A&M University
College Station, TX, USA
roozbeh@tamu.edu

Wendy Ham
Wharton School
University of Pennsylvania
Philadelphia, PA, USA
wham@wharton.upenn.edu

***Abstract*—We present a network study of the Santa Fe Institute Complex Systems Summer School (SFI CSSS) 2009. SFI CSSS 2009 participants consisted of complex systems researchers who were competitively selected to attend the month-long program. The international and interdisciplinary group spent four weeks together attending lectures, engaging in discussions, and collaborating on projects related to complex systems. We found that the CSSS 2009 network became increasingly decentralized and heterogeneous at the end of the program.**

***Index Terms*—Social Networks, Network Dynamics, Santa Fe Institute, Summer School**

I. Introduction

[5] describes a social network as "a set of people or groups of people with some pattern of contacts or interactions between them." Furthermore, a social network displays "a social structure made of individuals called 'nodes,' which are connected by one or more specific types of interdependency, such as friendship, kinship, financial exchange, dislike, sexual relationships, or relationships of beliefs or knowledge." Some examples of past studies on social network include friendships, business relationships, and academic collaborations.

In this paper, we present a social network analysis of a community of interdisciplinary researchers attending a month-long program organized by the Santa Fe Institute (SFI): the Complex Systems Summer School (CSSS). CSSS is an annual program that draws researchers from diverse disciplines who are selected through a competitive application process. CSSS participants spend four weeks together attending lectures, engaging in discussions, and collaborating on projects related to complex systems.

In 2009, approximately fourteen academic disciplines and nineteen nationalities were represented. CSSS 2009 participants primarily consisted of graduate students of all levels but also included several postdoctoral researchers, faculty, and industry practitioners. Due to the interdisciplinary and international nature of the group, almost none of the participants had known each other prior to the program. Thus, in CSSS 2009 a social network evolved virtually from scratch during the four weeks of the program, which provided a unique setting for a study of social network.

Throughout the entire program, all CSSS 2009 participants were encouraged to spontaneously form collaborative teams and work on one or more projects related to complex systems. In the first two weeks, the participants were free to explore each others research interests, whereas the last two weeks were dedicated more toward project-specific efforts. The participants pursued thirty-four projects by the end of the four weeks, all of which culminated in five-minute presentations and a poster session. All participants were strongly encouraged to submit working papers by September 1st. Beyond these short-term milestones, the participants had the option of expanding their summer projects into larger long-term projects. Nevertheless, it is fair to characterize CSSS 2009 projects as primarily exploratory.

Some of the main goals of CSSS are to promote interdisciplinary collaboration and to facilitate the cross-pollination of ideas. We are thus interested in understanding whether the patterns of evolution of the CSSS 2009 social network characteristics appear to support these goals. For example, does the CSSS 2009 network resemble the kind of network that allows ideas to flow freely? The high level of diversity in the CSSS 2009 participants background is especially relevant to this questiondoes interdisciplinary makeup lead to particular network characteristics? As a first step in this direction, we attempt to answer the following specific questions: i) What is the structure of the CSSS 2009 social interaction network? ii) How does this network change over time? iii) What factors might influence the evolution of the network? We apply network theory to provide a statistical characterization related to these questions as well as an understanding of the processes behind related network interactions [1].

II. Methods

We gathered data by distributing questionnaires at three different time points. We subsequently analyzed these data using statistical and network approaches. Our sample and methods are described in the subsections below.

A. Data collection

All CSSS 2009 participants were encouraged to fill in our social network questionnaires. As we previously describe, the participants consisted of graduate students (both Masters and PhD students), postdoctoral researchers, recent Masters and PhD graduates, industry practitioners from the SFI business network, one SFI staff, and two faculty members who were closely involved in all CSSS 2009 activities. The participants represented approximately fourteen disciplines and nineteen nationalities. Self-reported information describing the identity of the participants and project group composition is publicly available on the CSSS 2009 wiki.

We distributed questionnaires to a total of 64 participants at three different time points. The first questionnaires were distributed in the second week of the summer school, whereas the second and third questionnaires were distributed at the end of weeks three and four, respectively. The first survey included questions about individual attributes as well as network information. The second and third surveys focused only on network information. The attribute questions concern demographic information (nationality, gender, age, mother tongue, English fluency), dormitory location, areas of academic interest, academic institutions (undergraduate, graduate and postgraduateif applicable), and questions about social activity involvements. The network questions concern social interactions related to friendship and potential research collaborations. The first questionnaires were distributed on paper, whereas the second and third were distributed electronically on SurveyMonkey.com.

Our questionnaire design is based on the roster choice method [6]. Following this method, respondents were given a list of all other respondents (i.e., the roster) from which they selected people with whom they were friends and with whom they were interested in collaborating. Respondents were asked to indicate the strength of each relationship according to three levels: weak, moderate, and strong. We treat the resulting adjacency matrix as asymmetric in order to preserve the directional aspect of ties between respondents. In our current analysis, only strong ties are taken into consideration (binary value 1), whereas the weak and moderate ties are considered insignificant (binary value 0).

The response rate decreased over the survey period (perhaps due to fatigue from repetition and length of the survey): from 86% in the first survey, to 76% in the second, to 54% in the third. However, no bias was observed in the decreasing response. In order to be consistent, only those respondents are included in the analysis who replied to all the three surveys. We admit that while this improves the understanding of the evolution of their relationships across time, it excludes respondents whose relationships may have evolved in a manner different to those who replied to all surveys. Since the included sample of 27 respondents makes 43% of the original network, we assume it is meaningful for drawing conclusions about the general characteristics of the network.

B. Network Analysis

1) Inward and outward ties: We generate descriptive statistics for inward, outward, and all ties. The mean of inward/outward ties describes the proportion of existing inward/outward ties relative to all possible inward/outward ties. For example, a mean outward ties of 0.20 indicates that only 20% of all possible outward ties actually exist in the network. Network density describes the proportion of existing ties (both inward and outward) relative to all possible ties. Similarity describes the level of agreement among respondents in terms of with whom they form ties.

2) Graphical and qualitative analysis of clustering patterns: We use NetLogo to generate graphical representations of the network using data from the first survey. The purpose is to identify qualitatively any clustering patterns based on gender, academic discipline, and dormitory assignment.

3) Subgroups: We use Ucinet (Borgatti, Everett, and Freeman, 2002) and NetDraw (Borgatti, 2002) to identify subgroups in the network using k-core approach. As described by Hanneman and Riddle (2005): “k-core is a maximal group of actors, all of whom are connected to some number (k) of other members of the group The k-core approach [allows] actors to join the group if they are connected to k members, regardless of how many other members they may not be connected to.”

4) Centrality: We use Ucinet (Borgatti et al., 2002) to generate descriptive statistics of two centrality measures: degree (using Freeman’s approach) and betweenness. Even though degree and betweenness are generally used to assess the centrality of a node in the network, we focus on the interpretation of these measures for network-level centrality rather than actor-level centrality.

The saliency of degree as a measure of centrality is explained by Hanneman and Riddle (2005) as follows, “Actors who have more ties to other actors may be advantaged positions. Because they have many ties, they may have alternative ways to satisfy needs, and hence are less dependent on other individuals. Because they have many ties, they may have access to, and be able to call on more of the resources of the network as a whole. Because they have many ties, they are often third-parties and deal makers in exchanges among others, and are able to benefit from this brokerage. So, a very simple, but often very effective measure of an actor’s centrality and power potential is their degree.”

Because our network data are directed, we distinguish between in-degree and out-degree centrality. In-degree ties indicate prominence or prestige, whereas out-degree ties indicate influence (Hanneman and Riddle, 2005). The saliency of betweenness as a measure of centrality is explained by Garson (2009): “Betweenness measures how many paths pass through a node. Nodes high on betweenness have high opportunity to play gatekeeper, liaison, or broker roles. Betweenness is sometimes interpreted as ’influence’ or as degree of ’control of information’.”

III. Results and discussions

A. Outward/inward ties and reciprocity of ties

As shown in Table 1, the mean outward ties are similar to the mean inward ties. The disagreement (as indicated by standard deviation) between people over incoming ties, however, is larger and, surprisingly, grows toward the third survey. This is strange because it would be expected that as people get to know each other well they have a better understanding of the reciprocity of their relationships with others. It is possible, however, that because the final survey was conducted when the respondents had already left the program and thus were no longer influenced by the proximity of other respondents, they could more freely reflect on their ties with others.

B. Network density and similarity of ties

An important finding in our study is that while the density of the network increased between the first and second surveys, it decreased between the second and third surveys (Table 1). This fluctuation suggests that by the end of the program people had settled into stable ties (perhaps mostly dominated by ties from their work groups). This pattern is also indicated by the average sum of their connections, which increased from 5.26 to 6.11 between the first and second surveys only to return to 5.26 in the third survey. Furthermore, the persistent increase in the similarity of ties across respondents suggests that the respondents' approach in forming ties became more homogenous toward the end of the program.

C. Clustering patterns

As shown in Figure 2, we observe a moderate level of clustering based on gender. It is possible that this clustering is due to the overwhelming presence of males in the network. (Approximately sixty-four percent of our respondents shown in this figure are male.)

Figure 3 shows that people from the same disciplines (as indicated by the same-colored nodes) have some tendency to be located near each other in the network. For example, the grey nodes are adjacent to one another in the upper right portion of the graph, and the brown nodes are adjacent to one another in the lower left portion of the graph. This tendency is by no means overwhelming, however, as shown, for example, by the dispersed blue and green nodes. It is possible that this clustering tendency discipline-specific.

Figure 4 suggests some level of clustering according to dormitory assignments. For example the yellow nodes occupy a significant portion of the right side of the graph. Similarly, we observe some clustering of brown nodes in the lower portion of the graph.

D. Subgroups

Figures 5, 6 and 7 show the results of our k-core analyses for the first, second, and third surveys, respectively. Each color indicates a distinct subgroup, whereas the size of each node corresponds to its degree. As shown in Figure 5, two main subgroups are identifiable from the first survey. There is also a third subgroup, which consists of only one node (#18). Figure 6 shows again two main subgroups, but the second subgroup (indicated by the blue nodes) is much smaller than the first (red). Figure 7 shows seven subgroups, with one subgroup consisting of only one node (#12). Taken together, these results can be interpreted to mean that during the periods covered in the first and second surveys, CSSS participants engaged in exploratory social behavior, which peaked around the second survey. In this exploratory phase, the entire network is tightly connected with very few subgroups (especially around the time of the second survey). It is interesting to note that even as CSSS participants settled into work groups (around the second survey or week 3 of the program) the network connection pattern still indicates pervasive exploratory behavior. At the end of the program, however, as shown in Figure 7, CSSS participants appear to have settled into smaller subgroups. At this point it is unclear whether these subgroups are primarily social or influenced by project group interactions.

E. Centrality

1) Degree centrality: Tables 8, 9 and 10 show descriptive statistics of degree centrality from the first, second, and third surveys, respectively. First, we note that the OutDegree network centralization measure decreases a little from the first to the second survey (from 58.88% to 51.48%), then increases dramatically from the second to the third survey (from 51.48% to 82.84%). These results indicate that the outward ties became increasingly centralized at the end of the program, which suggests an unequal distribution of outward ties in the network at the end of the program (some actors had many more outward ties than others).

Second, we note that the InDegree network centralization measure shows the opposite pattern: it increases a little bit from the first to the second survey (from 18.93% to 19.53%) and decreases dramatically from the second to the third survey (from 19.53% to 10.95%). These results indicate that the inward ties became increasingly decentralized at the end of the program, meaning that most actors in the network probably had similar numbers of inward ties.

Third, we note that the OutDegree coefficient of variation (CV) decreases from the first to the second survey (from 85.13 to 68.17) and increases from the second to the third survey (from 68.17 to 107.64). CV is a measure of homogeneity in network ties (the lower the value the more homogeneous). So these results indicate that outward ties became more homogeneous from the first to the second survey and became very heterogeneous toward the end of the program. These results are consistent with the increasingly centralized outward ties at the end of the program; the structural positions of actors with regards to outward ties are heterogeneous at the end of the program, indicating unequal distribution of outward ties.

Taken together, all of the results above suggest that actors in the evolved version of the CSSS 2009 network may have similar levels of information inflow but varied levels of information outflow.

2) Betweenness centrality: Tables 11, 12 and 13 show descriptive statistics of betweenness centrality from the first, second, and third surveys, respectively. Consistent with the results of degree centrality analyses, the network became increasingly centralized at the end of the program: The centralization index increased from 16.48% (first survey) to 16.63% (second survey) to 26.36% (third survey). These values of centralization index are generally low, however.

	OutDegree	InDegree
Mean	5.26	5.26
Std Dev	4.48	1.71
Sum	142.00	142.00
Variance	20.04	2.93
Minimum	0.00	3.00
Maximum	20.00	10.00
Coefficient of Variation (CV)	85.13	32.57
Network Centralization (OutDegree)	=	58.88%
Network Centralization (InDegree)	=	18.93%

Fig. 8. Degree centrality: Survey I

	OutDegree	InDegree
Mean	6.11	6.11
Std Dev	4.17	2.08
Sum	165.00	165.00
Variance	17.36	4.32
Minimum	2.00	3.00
Maximum	19.00	11.00
Coefficient of Variation (CV)	68.17	34.02
Network Centralization (OutDegree)	=	51.48%
Network Centralization (InDegree)	=	19.53%

Fig. 9. Degree centrality: Survey II

	OutDegree	InDegree
Mean	5.26	5.26
Std Dev	5.66	1.67
Sum	142.00	142.00
Variance	32.04	2.78
Minimum	0.00	1.00
Maximum	26.00	8.00
Coefficient of Variation (CV)	107.64	31.74
Network Centralization (OutDegree)	=	82.84%
Network Centralization (InDegree)	=	10.95%

Fig. 10. Degree centrality: Survey III

	Betweenness	nBetweenness
Mean	32.93	5.07
Std Dev	35.55	5.47
Sum	889.00	136.77
Variance	1.263.78	29.91
Minimum	0.00	0.00
Maximum	136.07	20.93
Centralization Index	=	16.48%

Fig. 11. Betweenness centrality: Survey I

The standard deviation and variance decreased from the first to the second survey, and increased dramatically from the second to the third survey. These results are consistent with increased heterogeneity and decentralization in the network at the end of the program.

IV. Future directions

We have only scraped the surface of all the possible analyses that can be done with the data that we have collected. In future iterations of our study, we would like to take into consideration the strength of ties (weak, moderate, strong) rather than simplifying it into binary values. We also wish to link the friendship survey with the survey on collaboration network to show interactions between the two networks. Most practically, we would like to study the relationship between network positions of individuals and the quality of projects that they produce. Lastly, we hope that not too far in the future a similar study on CSSS social network can utilize advanced technology such as electronic sensors to gather far richer and more accurate data than we were able to gather in this study.

V. Acknowledgments

We thank the Santa Fe Institute and CSSS 2009 participants for their support of this project. We also thank Dr. Tom Carter and Dr. Nathan Eagle for their generous and helpful comments.

References

[1] Boccalettia, S.; Latorab, V.; Morenod, Y.; Chavezf , M. & D.-U. Hwanga (2006): Complex networks: Structure and dynamics. Physics Reports, 424:175 308.

[2] Borgatti, S. P. 2002. Netdraw network visualization. Harvard, MA: Analytic Technologies.

[3] Garson, G. D. 2009. http://faculty.chass.ncsu.edu/garson/PA765/networkanalysis.htm. Accessed on August 31st, 2009.

[4] Hanneman, R. A and Riddle, M. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside.

[5] Newman, M. E. J. (2003): The structure and function of complex networks. SIAM Review, 45(2):167256.

[6] Scott J., (2000). Social Network Analysis. Sage Publications: London.

	Betweenness	nBetweenness
Mean	26.07	4.01
Std Dev	27.30	4.201
Sum	704.00	108.31
Variance	745.55	17.65
Minimum	0.89	0.14
Maximum	130.13	20.02
Centralization Index	=	16.63%

Fig. 12. Betweenness centrality: Survey II

	Betweenness	nBetweenness
Mean	29.37	4.52
Std Dev	41.44	6.38
Sum	793.00	122.00
Variance	1.717.54	40.65
Minimum	0.00	0.00
Maximum	194.38	29.90
Centralization Index	=	26.36%

Fig. 13. Betweenness centrality: Survey III

Survey	Out			In			Density	Similarity
	Mean	Std Dev	Sum	Mean	Std Dev	Sum		
1	0.20	0.35	5.26	0.20	0.39	5.26	0.20	0.03
2	0.24	0.38	6.11	0.24	0.41	6.11	0.24	0.04
3	0.20	0.30	5.26	0.20	0.39	5.26	0.20	0.06

Fig. 1. Descriptive statistics of outward and inward ties, network density, and similarity of ties

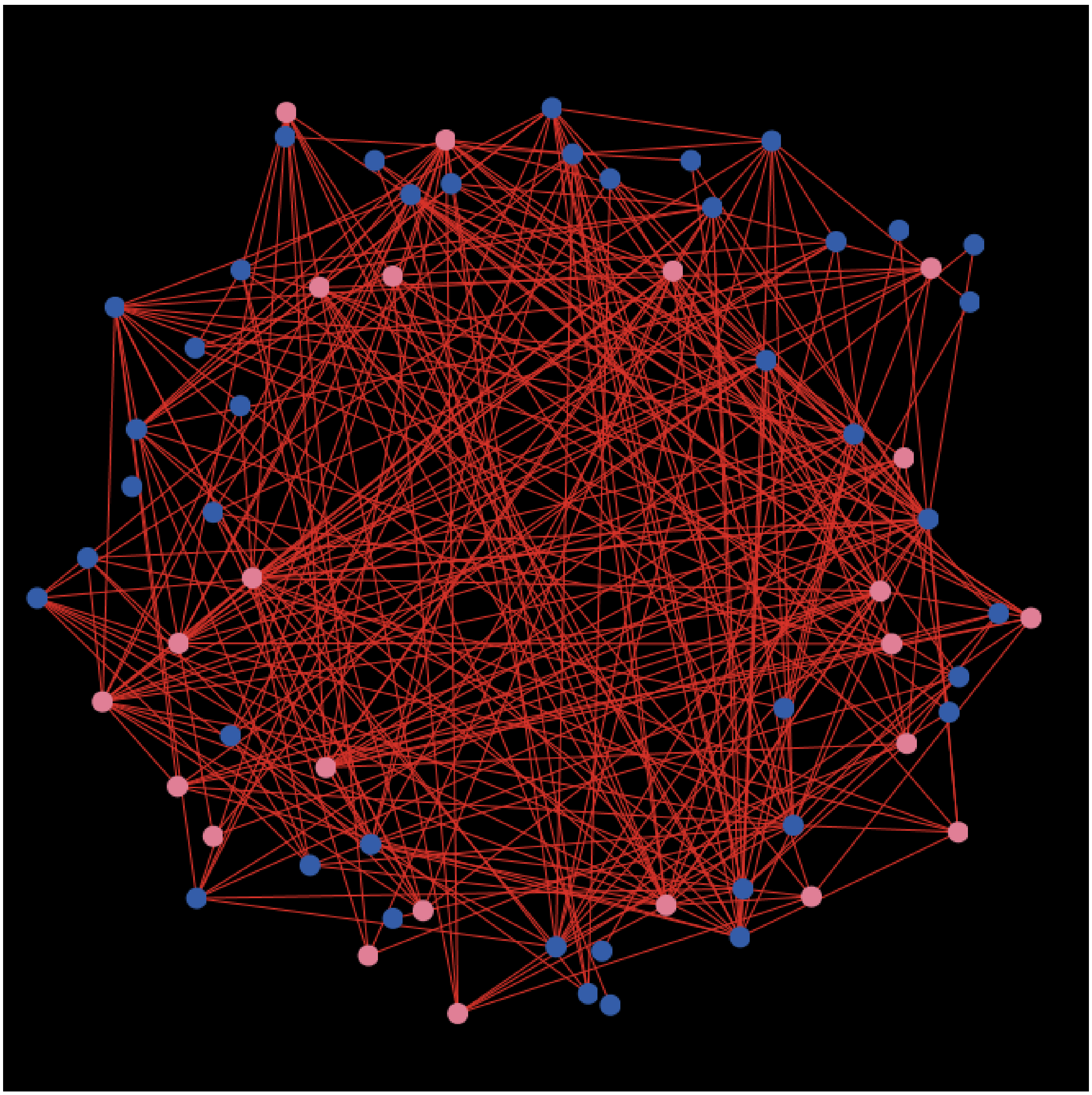

Fig. 2. Members in the network: Pink colors show females and blues show males

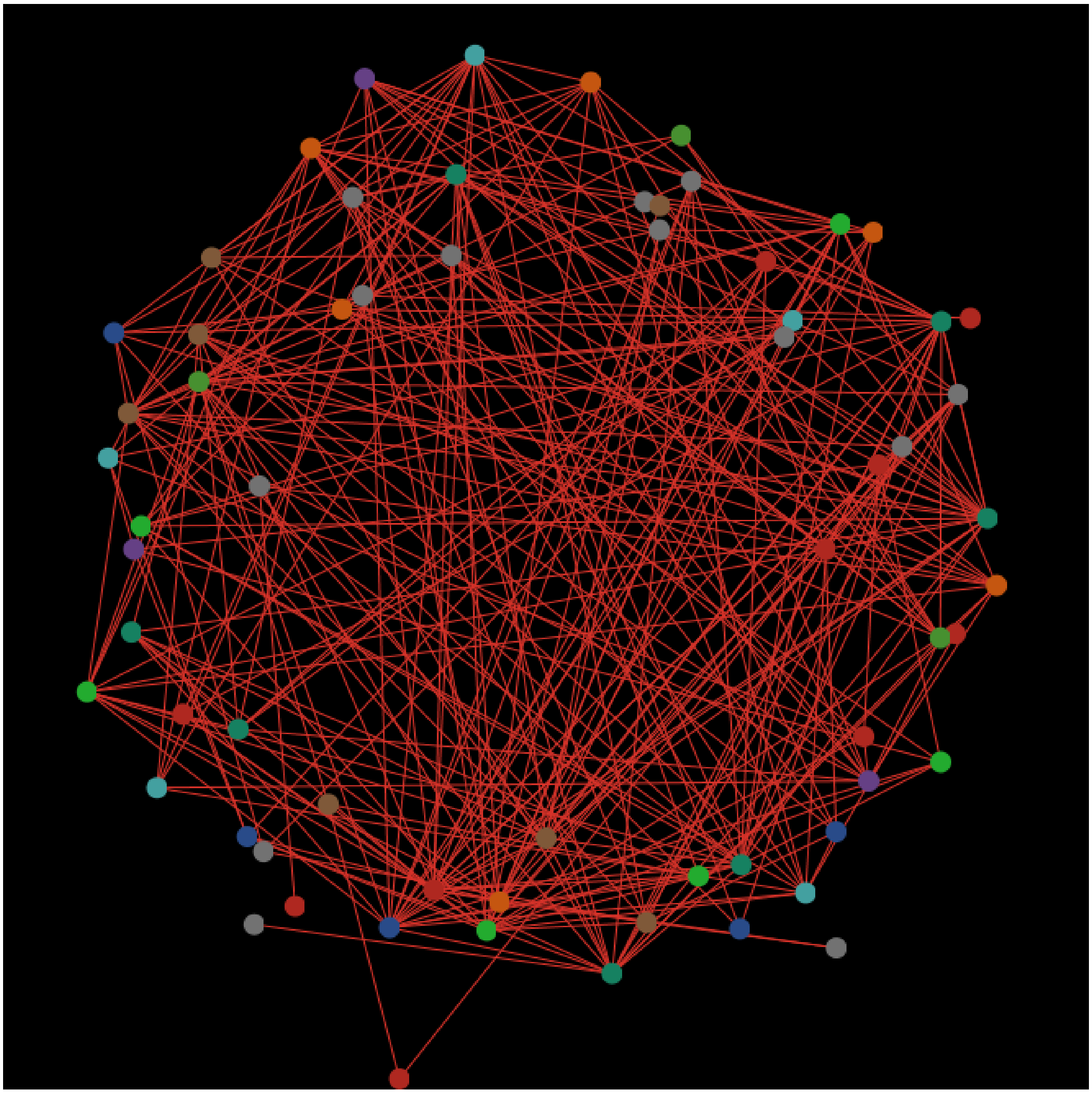

Fig. 3. Members in the network: Each color shows an area of research for the individual

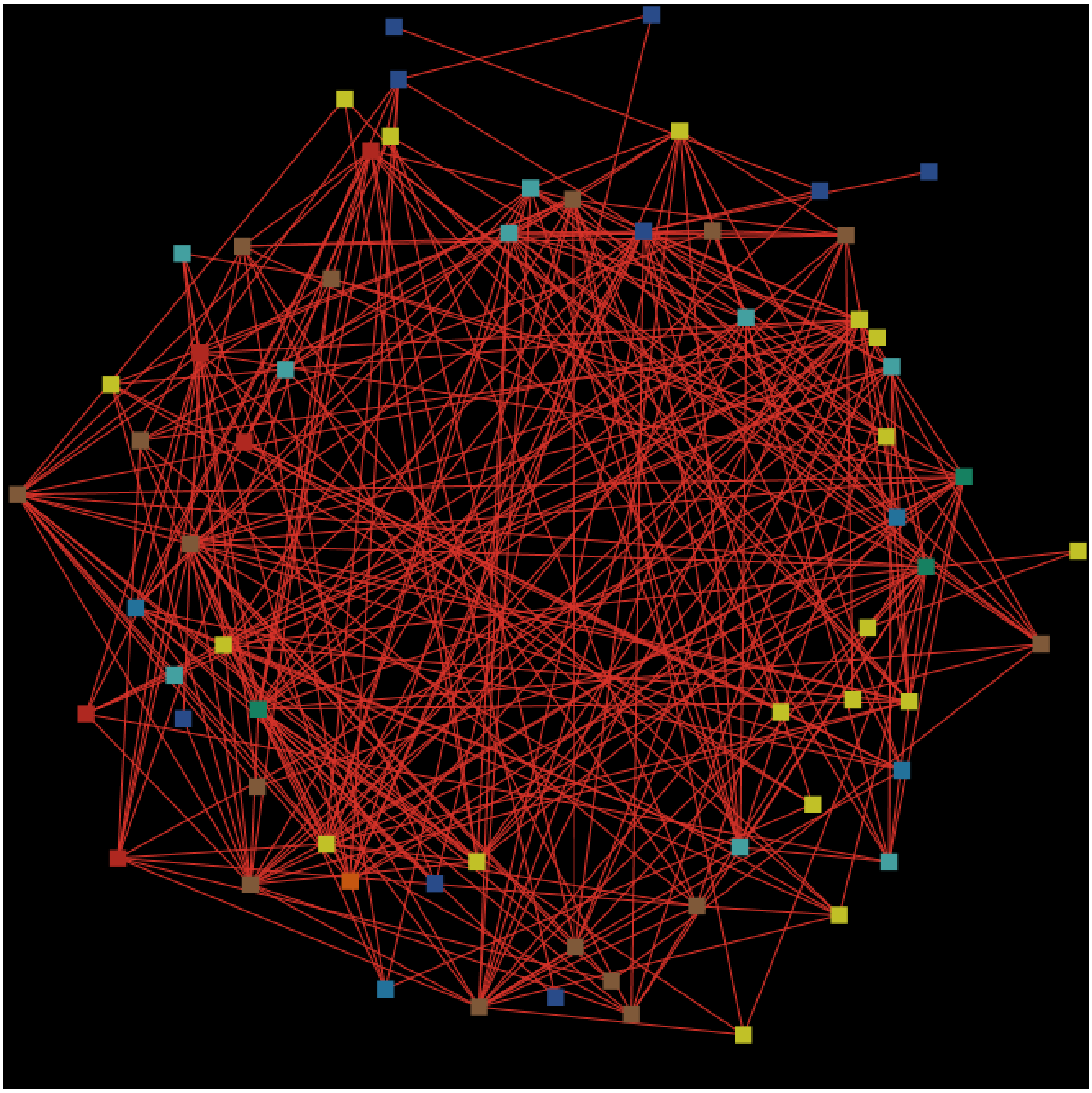

Fig. 4. Members in the network: Each color shows a specific dorm

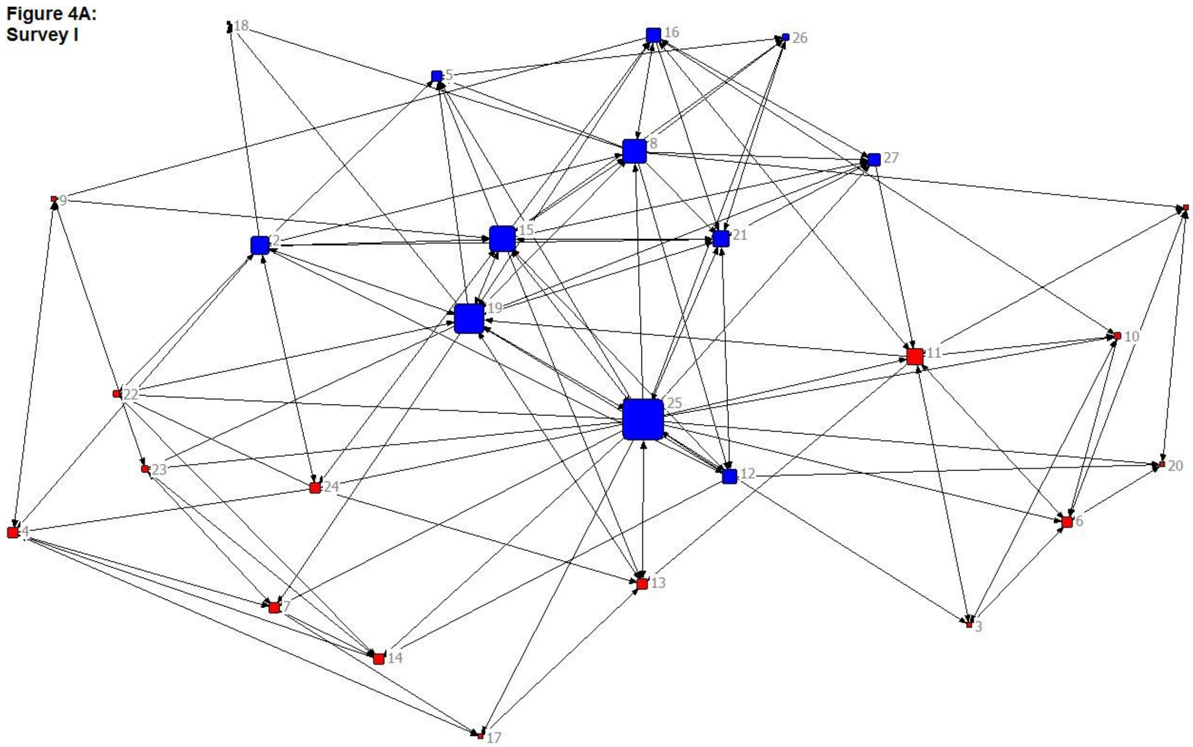

Fig. 5. Subgroups in CSSS 2009 network

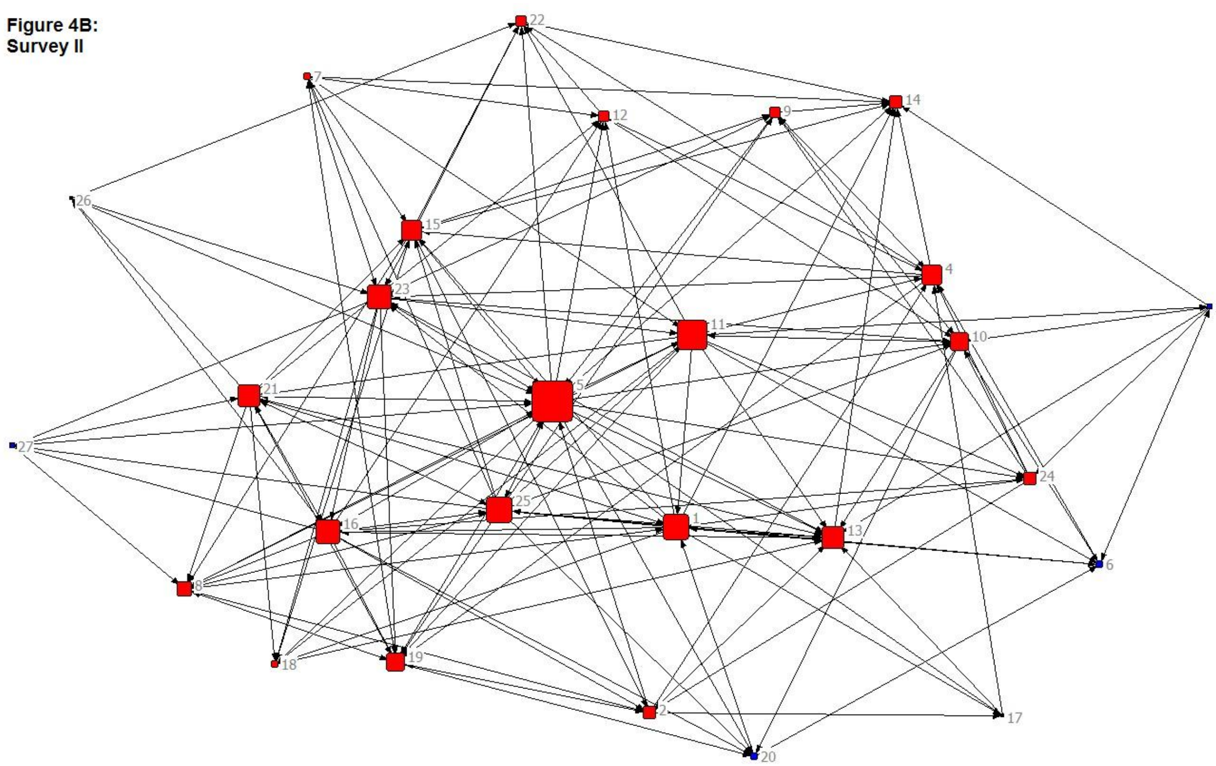

Fig. 6. Subgroups in CSSS 2009 network

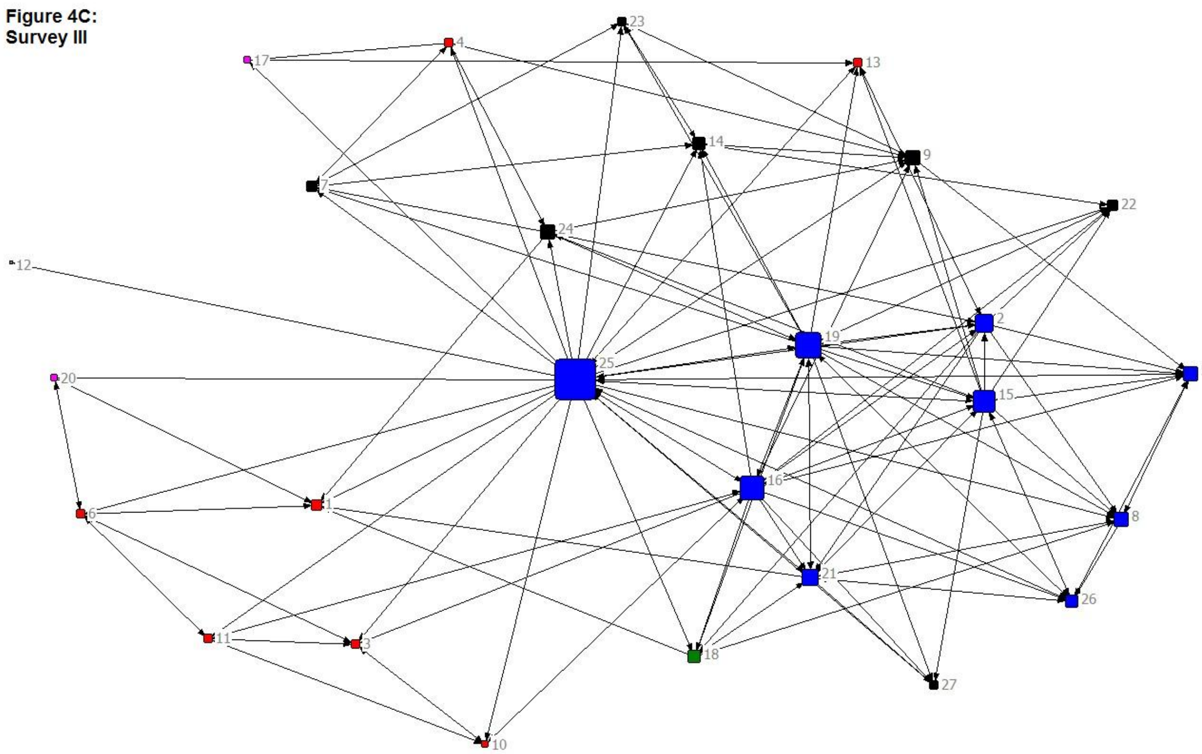

Fig. 7. Subgroups in CSSS 2009 network