DECONSTRUCTING CSSS 2009 SOCIAL NETWORK

ROOZBEH DANESHVAR
Department of Electrical Engineering, Texas A&M University
roozbeh@tamu.edu

WENDY HAM
The Wharton School, University of Pennsylvania
wham@wharton.upenn.edu

MARGRETH KEILER
Department of Geography and Regional Research, University of Vienna
margreth.keiler@univie.ac.at

MURAD MITHANI
The Lally School of Management and Technology, Rensselaer Polytechnic Institute
mitham@rpi.edu

Abstract—**We present a social network analysis of the Santa Fe Institute Complex Systems Summer School (CSSS) 2009. 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 collected network data at three different time points during the program to observe the evolution of the CSSS 2009 network structure. Our results suggest that: 1) The CSSS 2009 network was fluid: it appears that links could form and break quickly; 2) The network was relatively decentralized; 3) The network had relatively low average distance, low compactness, and high breadth; 4) The network was sparser in the middle of the program than at the beginning or at the end; 4) In general, CSSS 2009 participants appear to have mixed well; 5) Similarities in areas of study, however, appear to have influenced the formation of new links as well as the maintenance of potential long-term links.**

I. INTRODUCTION

In this paper, we present a social network analysis of a community of interdisciplinary researchers who attended a month-long program at the Santa Fe Institute (SFI): the Complex Systems Summer School (CSSS). CSSS is an annual program, and its participants typically come from diverse disciplines and are selected through a competitive application process. In four weeks, CSSS participants go through an intensive introduction to complex behavior in mathematical, physical living, and social systems. The participants spend a significant amount of time together attending lectures (four to five hours a day, five days a week), engaging in discussions, and collaborating on projects related to complex systems. Almost all participants live on-campus for the whole duration of the program, thereby providing plenty of opportunities to socialize in the off-hours.

In 2009, sixty-four participants represented approximately sixteen academic disciplines and twenty-three nationalities, which was in line with the usual interdisciplinary and international character of CSSS. CSSS 2009 participants primarily consisted of graduate students of all levels but also included several postdoctoral researchers, faculty, and industry practitioners. Almost none of the participants had known each other prior to the program. Thus, in CSSS 2009 a social network evolved virtually from scratch in four weeks, which provided a unique setting for a study of social network.

Throughout the program, all CSSS 2009 participants were encouraged to use the time outside of lecture hours to form collaborative teams and work on one or more projects related to topics in complex systems. In the first two weeks, the participants were free to explore each other's research interests, whereas the last two weeks were dedicated toward project-specific efforts. By the end of the program, CSSS 2009 participants collectively pursued thirty-four projects, all of which culminated in a series of five-minute presentations and a poster session. All participants were strongly encouraged to submit working papers by September 1st, 2009. Beyond these short-term milestones, all participants had the option of expanding their summer projects into larger long-term projects. Nevertheless, CSSS 2009 projects can be accurately characterized 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 CSSS 2009 social network display evolution patterns and other characteristics that would be consistent with these goals. The diverse backgrounds of CSSS 2009 participants are especially relevant—does interdisciplinary makeup lead to particular network characteristics? As a first step toward answering this question, we present our description and analysis of the following: i) detailed background of CSSS 2009 participants, ii) graphical snapshots of the CSSS 2009 network at three different time points, iii) evolution of degree measures, iv) evolution of distance measures, v) evolution of clustering coefficient, vi) evolution of homophily measures, and vii) correlations between attributes and network characteristics.

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 below.

A. Data collection

Our questionnaire design is based on the roster choice method (Scott, 2000). Following this method, respondents were given a list of all other respondents (i.e., the roster) from which to select with whom they were friends and were interested in collaborating. Respondents were asked to indicate the strength of each relationship: weak, moderate, or strong. 'Weak' was operationalized as 'conversing at least once since the start of CSSS 2009;' 'moderate' was operationalized as ‘conversing at least once a week;’ and ‘strong’ was operationalized as ‘conversing daily.’

We encouraged all sixty-four CSSS 2009 participants to fill out our social network questionnaires, which were distributed at three different time points. The first set of questionnaires was distributed at the beginning of the second week of the summer school, whereas the second and third sets were distributed at the end of the third and fourth weeks. The first survey included questions about individual attributes as well as relational information. The second and third surveys focused only on relational information. The first set of questionnaires was distributed on paper, whereas the second and third sets were distributed electronically using SurveyMonkey.com. The attribute questions pertain to demographic information (nationality, gender, age, native language, English fluency), on-campus dormitory location, areas of academic research, academic institutions (undergraduate, graduate and postdoctoral—if applicable), and questions about involvements in CSSS social activities. The relational questions pertain to friendship and potential research collaborations.

In this paper, we only present analysis of the friendship network (thereby excluding the research collaboration network), and we also exclude attribute data on: on-campus dormitory location, academic institution, and involvement in social activities. We exclude the research collaboration network because of our current limitation in time and space. We exclude on-campus dormitory location because CSSS 2009 participants rarely spent time in the dormitories other than to rest at night. We exclude academic institutions because they tend to be unique and therefore less meaningful for deducing patterns. Finally, we exclude involvement in social activities because it is beyond the scope of our current discussion. In spite of these justifications, however, we recognize the potential value of including all of the excluded attributes above in future analyses.

The response rate was 86% in the first survey, 76% in the second, and 54% in the third. We did not observe systematic biases in this decreasing pattern. Only those respondents who participated in all three surveys are included in our network analysis. Based on this criterion, twenty-seven respondents made it into our final network data set. One drawback of including only consistent respondents is that we may have inadvertently excluded respondents whose relationships evolved in a manner different to those included in the analysis. Nevertheless, given that the included sample of 27 respondents makes up 43% of the original network, we assume it is meaningful for drawing conclusions about the general characteristics of the network.

Even though the resulting adjacency matrix is asymmetric (which reflects the fact that perceptions of relationships are not always equal in a dyad), we choose to symmetrize it on the basis that a true link is one that is agreed upon by both members of a dyad. As such, we use the minimal (intersection) criteria to symmetrize the adjacency matrix. Furthermore, in the version of analysis that we present in this paper, only strong ties are taken into consideration (and assigned a binary value of 1), whereas weak and moderate ties are considered insignificant (binary value 0).

B. Network Analysis

To uncover patterns of network evolution, we use UCINET (Borgatti, Everett, and Freeman, 2002) and NetDraw (Borgatti, 2002) to generate the following analysis for each time point:

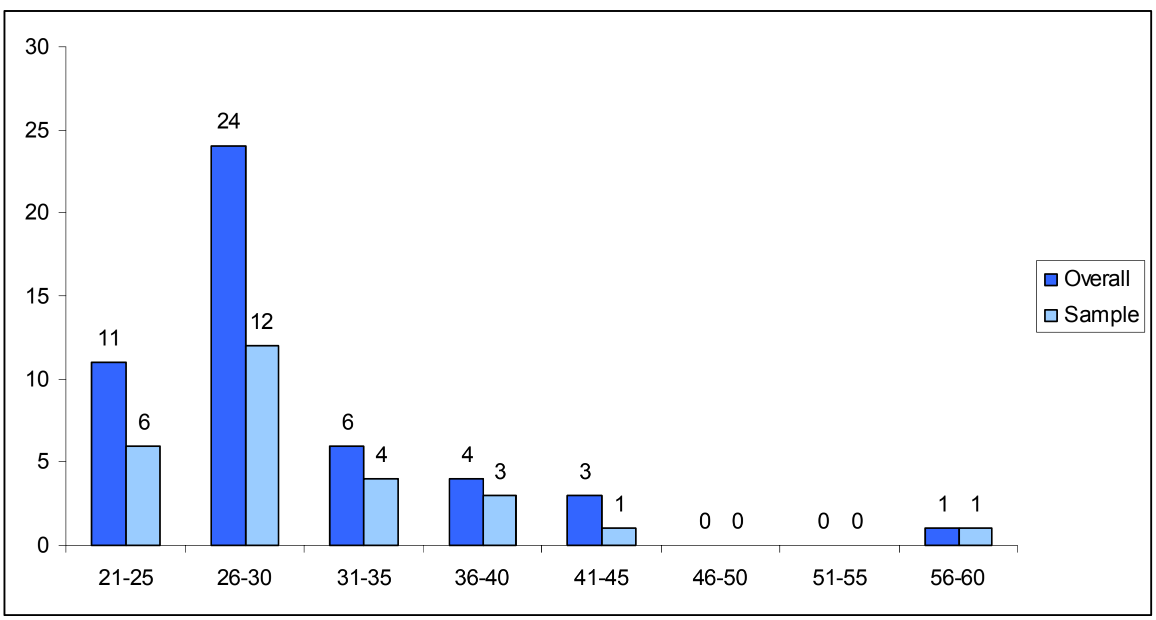
1) Network snapshots
2) Degree measures
3) Distance measures
4) Clustering coefficient
5) Homophily
6) Network and attribute data correlations

III. RESULTS AND DISCUSSIONS

A. Attribute data

1) Age: We received responses from 49 respondents (77% of all participants) with regards to age. The minimum, maximum, average, median, and standard deviation of the respondents' ages are 24, 59, 29.9, 27, and 6.7, respectively. In the sample of 27 respondents, the minimum, maximum, average, median, and standard deviation of the respondents' ages are 24, 59, 30.4, 27, and 7.7, respectively. The age distribution is shown in Figure 1.

Figure 1: Age distribution of CSSS 2009 participants



2) Gender: Of all 64 SFI CSSS 2009 participants, 41 (64%) are male and 23 (36%) are female. In the sample of 27 respondents, 16 (59.3%) are male and 11 are female (40.7%).

3) Nationality and country of residence: As shown in Table 1, all 64 CSSS 2009 participants represent 22 nationalities and 13 countries of residence. Most CSSS 2009 participants are US citizens (34 participants or 53.1%) and reside in the US (44 participants or 68.8%). There were four Canadian citizens, three Chinese, three Germans, two Italians, and two British. The rest of the countries are represented only by one citizen each. Four participants resided in the UK, three in Canada, three in Germany, and two in Singapore. The rest of the countries are represented by one resident each. In the sample of 27 respondents, sixteen nationalities and twelve countries of residence are represented.

Table 1: Countries represented in CSSS 2009

		OVERALL				SAMPLE			
	Country	Citizens		Residents		Citizens		Residents	
1	Australia	1	1.6%	1	1.6%	1	3.7%	1	3.7%
2	Austria	1	1.6%	1	1.6%	1	3.7%	1	3.7%
3	Belgium	0	0.0%	1	1.6%	0	0.0%	1	3.7%
4	Brazil	1	1.6%	0	0.0%	0	0.0%	0	0.0%
5	Bulgaria	1	1.6%	0	0.0%	1	3.7%	0	0.0%
6	Canada	4	6.3%	3	4.7%	2	7.4%	1	3.7%
7	China	3	4.7%	1	1.6%	2	7.4%	1	3.7%
8	Colombia	1	1.6%	0	0.0%	1	3.7%	0	0.0%
9	Germany	3	4.7%	3	4.7%	2	7.4%	3	11.1%
10	Hungary	1	1.6%	0	0.0%	1	3.7%	0	0.0%
11	India	1	1.6%	1	1.6%	0	0.0%	0	0.0%
12	Indonesia	1	1.6%	0	0.0%	1	3.7%	0	0.0%
13	Iran	1	1.6%	0	0.0%	1	3.7%	0	0.0%
14	Italy	2	3.1%	1	1.6%	2	7.4%	1	3.7%
15	Japan	1	1.6%	0	0.0%	0	0.0%	0	0.0%
16	Netherlands	0	0.0%	1	1.6%	0	0.0%	1	3.7%
17	Nigeria	1	1.6%	0	0.0%	0	0.0%	0	0.0%
18	Pakistan	1	1.6%	0	0.0%	1	3.7%	0	0.0%
19	Poland	1	1.6%	0	0.0%	1	3.7%	0	0.0%
20	Portugal	1	1.6%	0	0.0%	0	0.0%	0	0.0%
21	Russia	1	1.6%	0	0.0%	0	0.0%	0	0.0%
22	Singapore	0	0.0%	2	3.1%	0	0.0%	1	3.7%
23	Spain	1	1.6%	1	1.6%	1	3.7%	1	3.7%
24	UK	2	3.1%	4	6.3%	1	3.7%	3	11.1%
25	USA	34	53.1%	44	68.8%	8	29.6%	12	44.4%
	TOTAL	**64**	**100.0%**	**64**	**100.0%**	**27**	**100.0%**	**27**	**100.0%**

4) Language and English fluency: Language data were available from 53 respondents (83% of all participants). Of these respondents, thirty-one (58%) speak English as a native language. German and Chinese are the native languages of three participants each; Portuguese, Italian, and Spanish are the native languages of two participants each. Ten other native languages are spoken only by a single participant each. Thus, among CSSS 2009 participants English clearly does not have to compete with any other language. Furthermore, because non-English native languages tend to be spoken only by a single person or very few people, they were almost never used throughout the program. The average comfort level of using English among respondents was 2.8, where 3.0 was the highest comfort level. No respondent reported a comfort level of lower than two. It appears that language was not a barrier of interaction in the CSSS 2009 network. In the sample of 27 respondents, the average comfort level was 2.7.

5) Level of study: As shown in Table 2, among 62 respondents (95% of all participants), there were 50 (80.6%) graduate students, 4 (6.5%) postdoctoral researchers, 6 (9.7%) industry practitioners, 1 (1.6%) faculty, and 1 (1.6%) SFI staff. Of all graduate students in the data, we know that 17 (34.0%) were in year one or two of their studies, 8 (16.0%) were in year three or four of their studies, and 23 (46.0%) were in year five or beyond in their studies. (The other 2 are unknown.) Thus, it appears that CSSS 2009 graduate students were either looking for dissertation ideas or were near the end of their study (perhaps looking for postdoctoral directions)—not many were in between. In the sample of 27 respondents, 21 (77.8%) were graduate students, 4 (14.8%) were postdoctoral researchers, 1 (3.7%) was industry practitioner, and 1 (3.7%) was faculty. Among the graduate students in the sample, 14 (66.7%) were in year one or two, 2 (9.5%) were in year three or four, and 4 (19.0%) were in year five or beyond. Thus, with regards to level of study, there are appreciable differences between the composition of CSSS 2009 participants and the composition of our sample.

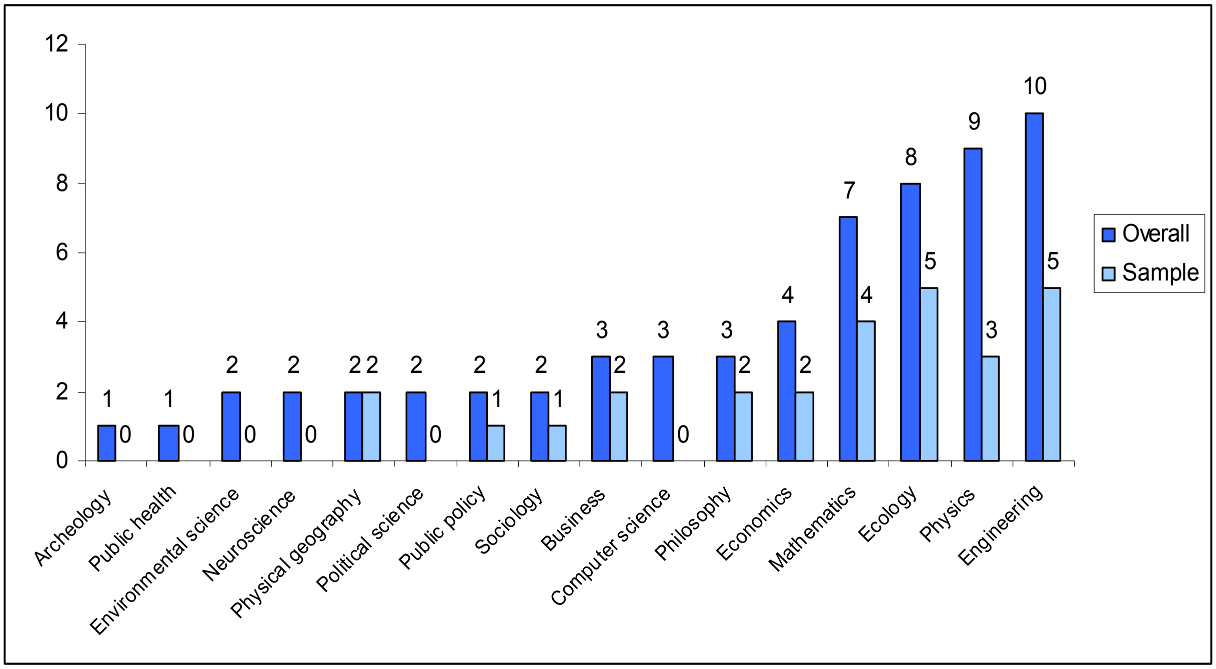
Table 2: Level of study of CSSS 2009 participants

Level of Study	**Overall**		**Sample**	
Graduate total	50	80.6%	21	77.8%
• Graduate year 1-2	17	34.0%	14	66.7%
• Graduate year 3-4	8	16.0%	2	9.5%
• Graduate year >5	23	46.0%	4	19.0%
• Graduate unknown	2	4.0%	1	3.7%
Postdoctoral	4	6.5%	4	14.8%
Industry	6	9.7%	1	3.7%
Faculty	1	1.6%	1	3.7%
Staff	1	1.6%	0	0.0%
TOTAL	**62**	**100.0%**	**27**	**100.0%**

6) Area of study: Among 61 respondents (95% of all participants), sixteen distinct areas of study were represented (Figure 2). Of these, 41 respondents (67%) were in natural science or engineering, and 20 (33%) were in humanities or social science. In the sample of 27 respondents, ten areas of study were represented. Of these respondents, 19 (70.4%) were in natural science or engineering, and 8 (29.6%) were in humanities or social science. Overall, most CSSS 2009 participants had strong quantitative background.

It is also interesting to note that approximately 60% of CSSS 2009 participants have switched academic fields. We define switching as having a degree in an area of study different than the current one. The areas of study must differ significantly. For example, a person who has an undergraduate degree in computer science who is now working in computational biology is not considered a switcher. In this case, even though the areas are different, the skill-sets involved are likely similar. Some examples of switchers in our data include a person who has an undergraduate degree in architecture who is now doing research in sociology, and another person who has an undergraduate degree in mechanical engineering who is now doing research in economics. In the sample of 27 respondents, we identified 9 (33.3%) switchers.

Figure 2: Areas of study represented in CSSS 2009



B. Network analysis

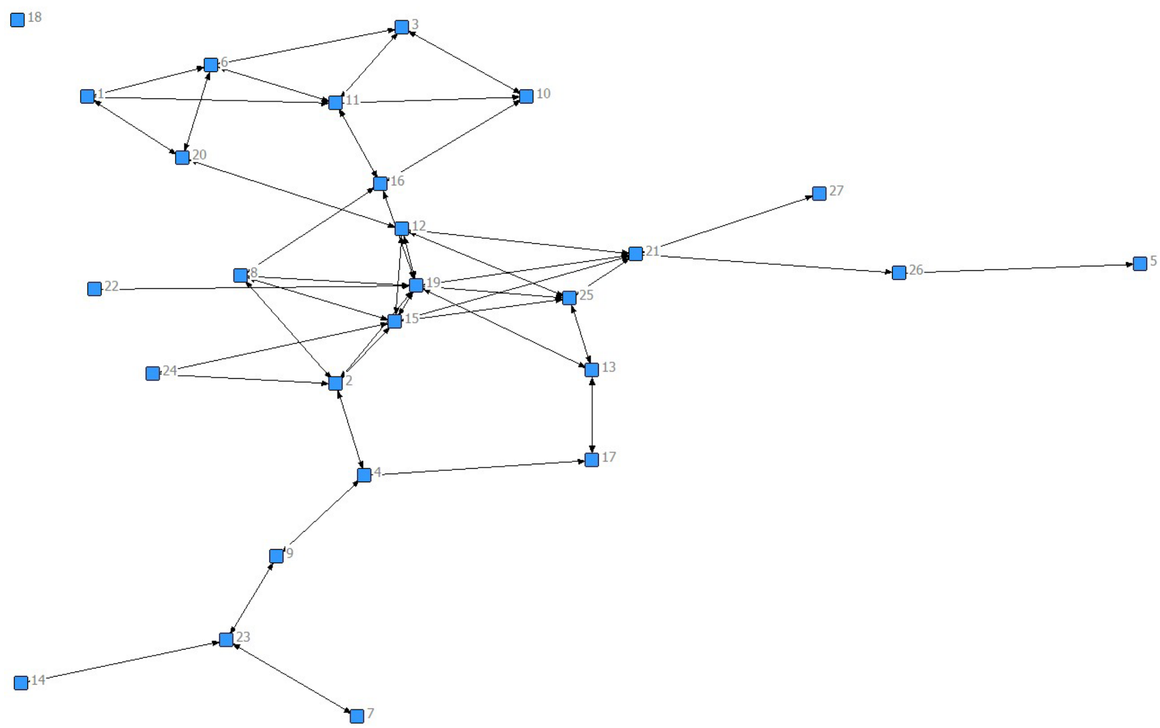
1) Network snapshots: As shown in Figure 3, the first network is almost fully connected; it consists of two components, one of which is an isolate. The second network consists of nine components, seven of which are isolates, and one of which is a two-node network. The third network consists of six components, four of which are isolates. The second network is noticeably sparser than the first and third networks.

It is interesting to note that even though these networks were separated by only one week (approximately), they appear significantly different. For example, subject #5 was a peripheral actor in the first network but became one of the obviously central actor in the second network. Conversely, subject #15 was quite central in the first network but became peripheral in the second network. Subject #12 was also quite central in the first network but became an isolate in the second network.

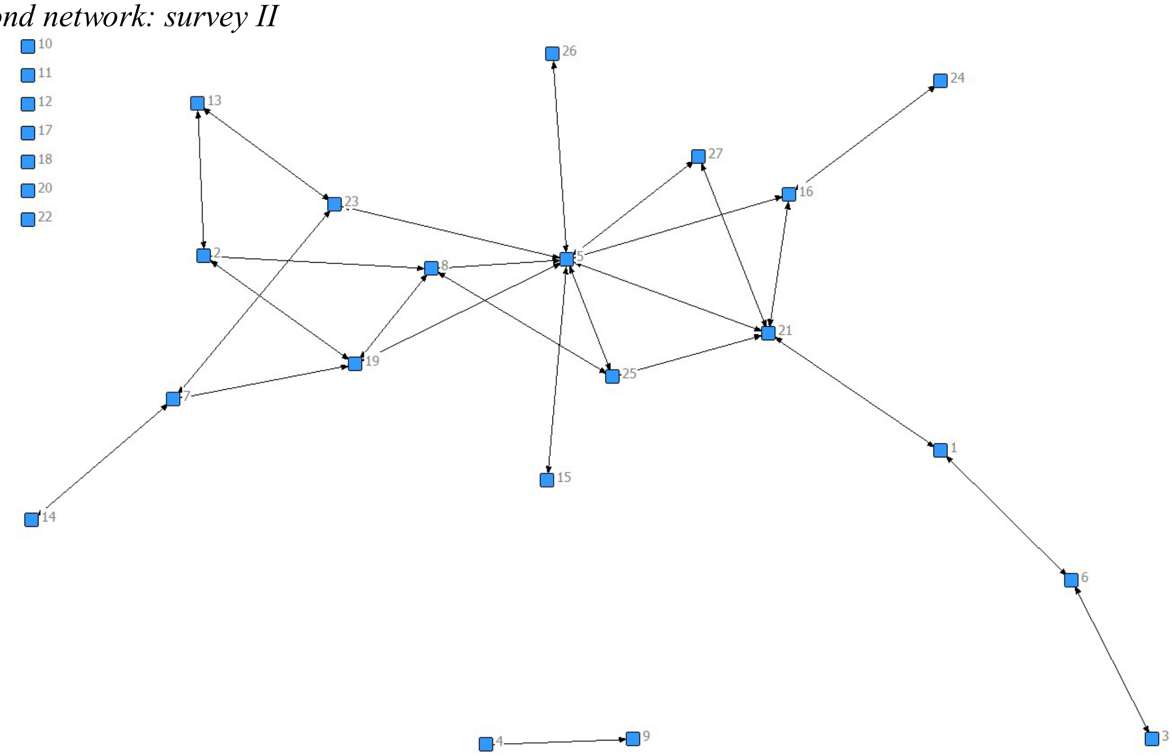
Taken together, these three snapshots suggest that the CSSS 2009 network was fluid: links could form and break rather quickly. The density of the first network is perhaps indicative of exploratory social behavior: up until the beginning of the second week CSSS 2009 participants were still trying to get to know everyone. The second and third networks, on the other hand, perhaps indicate the beginning of longer-term friendships.

Figure 3: CSSS 2009 network snapshots

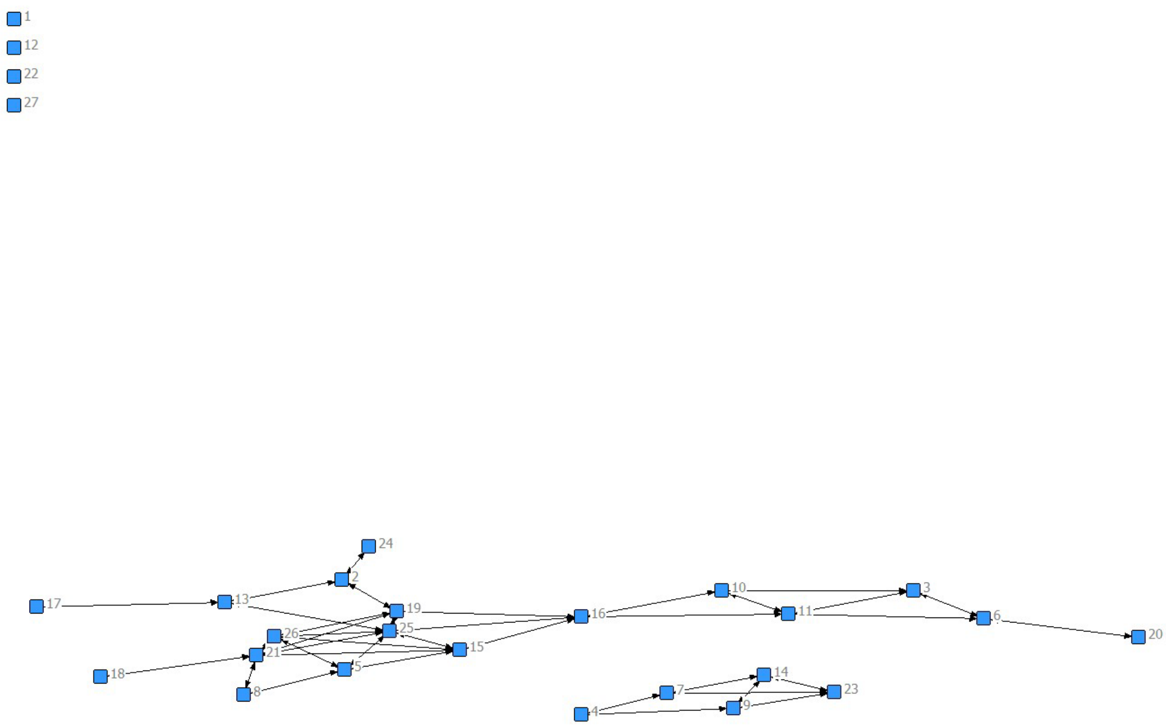
First network: survey I



Second network: survey II



Third network: survey III



2) Degree measures: We calculate descriptive statistics of degrees, as well as network centralization based on degree centrality. The CSSS 2009 network was relatively decentralized. The results are shown in Table 4.

Table 4: Degree measures

	First network	Second network	Third network
Mean	3.26	1.93	2.81
Minimum	0	0	0
Maximum	9	9	7
St. Dev	2.03	1.98	1.89
Sum	88	52	76
Network centralization	23.85%	29.38%	17.38%

3) Distance measures: We calculate average distance among reachable pairs, distance-based cohesion ("compactness"), and distance-weighted fragmentation ("breadth"). Distance-based cohesion ("compactness") is basically the sum of reciprocal distances, with the convention that the reciprocal of infinite or undefined distance is 0; it is normalized by dividing by the number of directed pairs (Borgatti, 2006). Distance-weighted fragmentation ("breadth") is defined as the average reciprocal distance among nodes after removal of a given node; its value is 1 when all nodes are distance 1 from each other (i.e., a complete graph), and 0 when all nodes are isolates. Intermediate values measure the extent to which the presence of a node tends to reduce distances in the network (Borgatti and Everett, 2006).

The results are shown in Table 5. Average distance decreased over time, although interestingly distance-based cohesion also decreased (perhaps due to the increase in the number of isolates). Distance-weighted fragmentation increased from the first to the second network and dropped slightly in the third network.

Table 5: Distance measures

	First network	Second network	Third network
Average distance (among reachable pairs)	3.372	2.721	2.669
Distance-based cohesion ("compactness")	0.378	0.208	0.234
Distance-weighted fragmentation ("breadth")	0.622	0.792	0.766

4) Clustering coefficient: We calculate weighted overall graph clustering coefficients for all three networks. The clustering coefficients suggest that the first and third networks are more cohesive than the second network. The results are summarized in Table 5 below.

Table 6: Weighted overall graph clustering coefficient

First network	0.426
Second network	0.234
Third network	0.436

5) Homophily: We measure the extent to which each subject has ties with others with the same attributes. Given that we are interested in how the interdisciplinary character of CSSS 2009 may influence social ties, we look at the attribute data that differentiate between subjects whose areas of study are in natural science or engineering and those in social science or humanities. We assigned a dummy variable of 0 to those in social science or humanities, and 1 to those in natural science or engineering. The homophily measure is binary: It is 1 if a link exists between a pair with the same attribute and 0 otherwise. The average homophily is 0.64 for the first network, 0.66 for the second, and 0.56 for the third network. As such, there is a decreasing pattern over time, although these differences are not statistically significant according to two-tailed t-test.

6) Network and attribute correlation: We compute pairwise correlations for the following variables (1-9 are network variables, and the rest are attributes):

1. Degree centrality in the first network
2. Degree centrality in the second network
3. Degree centrality in the third network
4. Clustering coefficient in the first network
5. Clustering coefficient in the second network
6. Clustering coefficient in the third network
7. Homophily (based on area of study as explained in point 5 above) in the first network
8. Homophily in the second network
9. Homophily in the third network
10. Age
11. Age group (1 = 21-25, 2 = 26-30, … , 8 = 56-60)
12. Gender (0 = male, 1 = female)
13. Nationality (arbitrarily assigned numeric ID for each nationality)

14. Country of residence arbitrarily assigned numeric ID for each country)
15. English fluency (2 or 3)
16. Whether English is a native language (0 = no, 1 = yes)
17. Level of study (1 = graduate, 2 = postdoctoral, 3 = industry, 4 = faculty)
18. Year of study (1 = years 1 and 2, 2 = years 3 or 4, 3 = years 5 and beyond)
19. Area of study (arbitrarily assigned numeric ID for each area of study)
20. Whether the subject switches field (0 = no, 1 = yes)
21. Whether the area of study is in natural science/engineering or social science/humanities (0 = social science/humanities, 1 = natural science/engineering)

The results are shown in Table 7.

Table 7: Correlation matrix of network and attribute data

	degreei	degreeii	degr~iii	clust~gi	clus~gii	clus~iii	homop~li
degreei	1.0000						
degreeii	0.1615	1.0000					
degreeiii	0.5157*	0.4426*	1.0000				
clusteringi	0.1557	0.0521	-0.1199	1.0000			
clusteringii	0.0821	-0.0566	0.0200	0.4470	1.0000		
clusteri~iii	-0.1298	-0.1899	0.3060	0.0308	-0.1845	1.0000	
homophily~li	-0.1332	0.0623	-0.2078	0.0905	-0.0718	-0.2564	1.0000
homophil~lii	-0.1309	-0.1846	-0.2564	-0.1306	-0.4993	-0.0717	0.2991
homophil~iii	0.0147	-0.1692	-0.2563	0.0886	-0.2352	-0.4101	0.6448*
age	0.3088	0.1853	0.3460	0.2218	0.2376	0.0934	0.1372
agegroup	0.2754	0.1492	0.3457	0.1907	0.2146	0.1376	0.1265
gender	-0.0316	-0.1974	-0.2383	-0.0851	-0.2959	-0.5001*	0.0730
nationality	-0.0055	0.1100	0.0314	0.0005	-0.3178	0.0837	0.0606
residence	-0.1279	0.0852	-0.0378	-0.3485	-0.1128	0.3477	0.0751
english	-0.3967*	0.0986	-0.1067	0.0854	-0.3137	0.2582	0.0556
native	-0.2439	0.2900	-0.0420	0.0983	-0.5467	0.1047	0.3442
levelofstudy	0.3464	-0.1126	0.2090	0.2969	0.3878	0.0123	0.0521
year	-0.3645	0.2349	0.0929	0.1208	-0.1606	0.0999	-0.0064
areaofstudy	-0.2367	-0.1389	-0.2224	-0.0054	0.0642	0.0737	0.2249
switcher	-0.0516	0.2249	0.2360	-0.2187	-0.1323	-0.0190	0.3053
natural	0.1229	-0.3520	-0.1497	0.1499	0.2216	-0.2401	0.3192

	homo~lii	homo~iii	age	agegroup	gender	nation~y	reside~e
homophil~lii	1.0000						
homophil~iii	0.1958	1.0000					
age	-0.4308	0.0242	1.0000				
agegroup	-0.4633*	0.0626	0.9873*	1.0000			
gender	0.0219	0.1789	-0.1584	-0.1657	1.0000		
nationality	0.1000	0.0446	0.4245*	0.3606	-0.1066	1.0000	
residence	0.1016	-0.1520	0.2732	0.2095	-0.2072	0.6958*	1.0000
english	0.2952	0.0715	0.1347	0.1542	-0.1223	0.2775	0.3091
native	0.3967	0.1242	-0.1139	-0.1122	-0.0806	0.4216*	0.2099
levelofstudy	-0.4385	0.0760	0.8370*	0.8116*	-0.0698	0.2975	0.1166
year	-0.2536	0.0909	0.2316	0.3228	-0.4551*	0.2553	-0.0842
areaofstudy	0.4726*	0.3555	-0.0669	-0.0896	-0.4617*	0.3591	0.2917
switcher	0.1926	-0.1690	0.3326	0.2968	0.2132	0.3228	0.3229
natural	-0.0272	0.6864*	0.0811	0.1001	0.0428	0.0834	-0.1395

	english	native	levelo~y	year	areaof~y	switcher	natural
english	1.0000						
native	0.3468	1.0000					
levelofstudy	-0.0375	-0.2474	1.0000				
year	0.3101	0.2707	.*	1.0000			
areaofstudy	0.2887	0.1738	-0.0158	0.2884	1.0000		
switcher	0.1147	0.1890	0.0000	-0.0650	-0.2178	1.0000	
natural	-0.0658	-0.0434	0.3004	0.1266	0.5135*	-0.4015*	1.0000

** = significant at p < 0.05 level*

In general, significant correlations are sporadic, which suggests that there was no systematic bias in the ways in which CSSS 2009 participants socialized with each other. Nevertheless, we discuss each significant correlation in Table 8 below:

Table 8: Significant correlations

No.	Variable 1	Variable 2	Correlation	Interpretation
1	Degree (first network)	Degree (third network)	0.5157	This correlation and correlation #2 below together suggest that a subset of subjects with high degrees in the first network drop their links in the second network but subsequently renew them in the third network.
2	Degree (second network)	Degree (third network)	0.4426	(See above)
3	Homophily (first network)	Homophily (third network)	0.6448	This suggests that new links (which were formed at the start of the program) and links that are likely to persist beyond CSSS 2009 were both influenced by similarities in areas of study.
4	Gender	Clustering coefficient (third network)	-0.5001	Being female is negatively correlated with having a high clustering coefficient in the third network. This suggests that female subjects are more likely to span cliques in the third network.
5	English fluency	Degree (first network)	-0.3967	This suggests that subjects who are fluent in English form links early on (in the first network), but subjects who are less fluent appear to catch up later on (in the second and third networks).
6	Age group	Homophily (second network)	-0.4633	This suggests that older subjects are more likely to form links with others who from different areas of study.
7	Age group	Age	0.9873	Obvious
8	Nationality	Age	0.4245	Because each nationality is assigned an arbitrary numeric ID, this correlation is not meaningful.
9	Country of residence	Nationality	0.6958	This suggests that CSSS participants tend to reside in their countries of citizenship.
10	Native language is English	Nationality	0.4216	Obvious
11	Level of study	Age	0.8370	Obvious
12	Level of study	Age group	0.8116	Obvious
13	Year of study	Gender	-0.4551	This suggests that CSSS participants who are at an advanced level of their studies tend to be male.
14	Area of study	Homophily (second network)	0.4726	Because each area of study is assigned an arbitrary numeric ID, this correlation is not meaningful.
15	Area of study	Gender	-0.4617	Because each area of study is assigned an arbitrary numeric ID, this correlation is not meaningful.
16	Natural science/engineering or social science/humanities	Homophily (third network)	0.6864	This suggests that subjects from natural science/engineering tend to maintain links with others who are also from natural science/engineering in the third network. This correlation is consistent with correlation #3 above.
17	Natural science/engineering or social science/humanities	Gender	0.5135	This suggests that subjects from natural science/engineering tend to be female.
18	Natural science/engineering or social science/humanities	Nationality	-0.4015	Because each nationality is assigned an arbitrary numeric ID, this correlation is not meaningful.

IV. CONCLUSION AND FUTURE DIRECTIONS

We have only scraped the surface of all the possible analyses that can be performed using the data we have collected. The nature of the work presented in this paper is exploratory—our primary aim is to present the CSSS 2009 network characteristics and to capture any discernible patterns in its evolution throughout the course of program.

Taken together, our results suggest that the CSSS 2009 network was fluid. The appreciable structural differences between the first, second, and third networks indicate that links could form and break quickly. Furthermore, CSSS 2009 network was relatively decentralized and had relatively low average distance, low compactness, and high breadth. The network became sparse approximately halfway through the program, but regained its original density toward the end of the program. In general, CSSS 2009 participants appear to have mixed well despite their diverse backgrounds, with the exception that similarities in areas of study appear to have influenced the formation of new links at the beginning of the program as well as the maintenance of potential long-term links at the end of the program.

Future iterations this study should: i) take into consideration the strength of ties (weak, moderate, strong) rather than simplifying it into binary values, ii) link the friendship network with the collaboration network to investigate any interactions between the two networks, and iii) study the relationship between network positions of CSSS 2009 participants and the quality of their projects. Lastly, we hope that in the near future a similar network study can utilize more sophisticated data collection tools, such as electronic sensors, to gather far richer and more accurate information about network dynamics than we were able to gather in this study.

V. ACKNOWLEDGMENTS

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

VI. REFERENCES

Borgatti, S. P. 2002. *Netdraw network visualization*. Harvard, MA: Analytic Technologies.

Borgatti, S. P. 2006. Identifying sets of key players in a social network. *Computational and Mathematical Organization Theory*, *12*:21-34.

Borgatti, S. P., Everett, M. G. and Freeman, L. C. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Borgatti, S. P., and Everett, M. G. 2006. A graph-theoretic perspective on centrality. *Social Network, 28*:466-484.

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

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

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Review, 45(2)*:167–256.

Scott J. 2000. *Social network analysis*. Sage Publications: London.