V.R.F. Series

No.498
Mar 2017

Structural Analysis of Networks in Indian Schools

Pablo Lucas

日本貿易振興機構 アジア経済研究所

INSTITUTE OF DEVELOPING ECONOMIES, JAPAN EXTERNAL TRADE ORGANIZATION

Structural Analysis of Networks in Indian Schools

Pablo Lucas

Acknowledgements

I would like to express my genuine gratitude for the opportunity to work collaboratively in this research project via the support given by the Visiting Research Fellow program, hosting me at the Institute of Developing Economies, Japan External Trade Organization (henceforth IDE-JETRO) and my counterpart Dr. Seiro Ito. His exceptional attention to provide constructive feedback is very much welcome, as this encourages improving the overall research process and project.

I would also like to thank the very useful discussions held at Tokyo University with project members: Dr. Nobuhiko Fuwa (Waseda University), Dr. Takashi Kurosaki (Hitotsubashi University), Dr. Yasuyuki Sawada (Tokyo University) and Dr. Seiro Ito (IDE-JETRO). Their willingness to collaborate boosts research as a whole and it is appreciated to work with such collegial and accomplished academics.

I am also grateful for the useful discussions held at IDE-JETRO with other research professionals, namely Dr.Yuri Sato, Dr. Yuko Tsujita, Dr. Tomohiro Machikita, Dr. Sonam Tobgyal (Bhutan Gross National Happiness Commission).

I would also like to thank the administrative support from Ms. Atsuko Hirakata, Ms. Manda Kumi, Mr. Takeo Masuda and Mr. Rihito Noritake. Their dedication and efficiency throughout and after the fellowship is much appreciated. That helped me to overcome various needs, ranging from the necessary paperwork through adapting to my needs of special software and hardware requirements.

I consider the research period at IDE-JETRO in Chiba an excellent experience, both personally and professionally. I look forward to our follow-up developments. Many thanks indeed for all your support, professionalism and cordiality in Japan!

Dr Pablo Lucas
December 2015

Contents

List of Tables

List of Figures

Abstract

This report complements the research project entitled "Values Affirmation Exercise in Indian Schools" (henceforth VAEIS), which has the purpose of assessing whether students realise their full learning potential through an experiment of respectfully asserting individual values aiming to influence academic achievement.

In this document there is an initial description of the structure of student networks according to their perceptions, as recorded by the administered surveys. Focus is given to discuss different cross-sectional student networks through the application of social network analysis (henceforth SNA) techniques to the data.

At the end of the VAEI project, a summary will be distributed to parents, teachers, officials, and academics. The ultimate aim of this collaboration is to publish results, academically and in policy-relevant venues. This report serves as a first description of the collected network structures and observed social dynamics.

1. Introduction

This research report complements the research project entitled "Values Affirmation Exercise in Indian Schools" (henceforth VAEIS), funded by the Japanese Government's Ministry of Education, supporting the Japan-based project members mentioned in the acknowledgements section. The support received from IDE-JETRO via the Visiting Research Fellow (henceforth VRF) program enabled to further the collaboration regarding student networks in the VAEIS project.

The purpose of the VAEIS project is to assess whether students are realising their learning potential at their fullest through a simple experiment of respectfully asserting the students' individual values as a potential way to help them to better achieve academically [VAEIS, 2015]. The objective of this research report is to initially describe the structure of student networks according to their perceptions, as recorded by the administered surveys. Thus the focus is the discussion of different cross-sectional student networks through the application of social network analysis (henceforth SNA) techniques to the collected dataset.

In this study, SNA allows describing the cross-sectional structure that result from likely social patterns, including structural properties such as measures of centrality (e.g. concepts of closeness, betweenness and eigenvector centrality) and others (e.g. concepts of transitivity, components, diameter and reciprocity)[1].

[1] The definitions of these and other concepts can be found in section 3.1 (from page 22).

The ongoing VAEI project will provide further data regarding the student learning processes after school hours, their relationship with achievements and impact of exercises on academic achievements. Due to the lack of precision, qualitative interpretations of network structures can be limited in contributing to understand which factor(s) indeed influence most the observed structure. An SNA quantitative approach can facilitate this assessment and description through more precise descriptions and tests of which combination of different network regularities can sustain observed structures rooted in social interaction [Lusher et al., 2012].

Therefore one could go a step further and test which potentially conflicting explanations fit best a given observed network dataset and infer its relationship to other outcomes. With ongoing developments in network model specification and estimation, it is possible to probe into the structure of networks with increasing precision. This requires developing models that account for the observable characteristics (i.e. network structures and possibly the attributes of individuals). This has been deemed future work [2] due to the current dataset limitations [3].

At the end of the VAEI project, a summary will be distributed to parents, teachers, officials, and academics. The ultimate aim of the collaboration is to publish results. This report serves as a first description of the collected network structures and social dynamics at the five surveyed schools and is structured as follows:

- section 2: "Research Design" (page 9) contains a discussion of aspects about the survey administrationand and time spent preparing the data for SNA;
- section 3: "Research Findings" (page 21) is focused on discussing the network composition, properties and structures in schools, across activities;
- section 4: "Final Considerations" (page 47) onwards concludes the report.

[2] Discussed in greater detail in section 4.1: "Future Work" (page 48).
[3] Described in greater detail in section 2.3: "Processing The Working Dataset" (page 14).

2. Research Design

This section is dedicated to explain and discuss how this research project has been designed and what has been carried out, research-wise, up to the date in which this research report has been last updated (i.e. 15th December 2015).

The VAEIS project incorporates an experiment in which the student's individual values are asserted in an attempt to help them to improve academic performance. For this reason there is a detailed experiment protocol for the field manager, including sections about what to do: (I) before going to a school, (II) at the school, (III) at the classroom, (IV) before leaving a school and (V) after leaving a school.

With the agreements from schools, governments and student carers, the students have been guided by their teachers throughout the data collection process. Teachers were trained by VAEIS project members to announce that participating students should provide written responses to questions that have been designed to allow understanding more about their ideas, beliefs, life in general and networks.[1]

The next sections briefly describe the research study, how data collection has been done and issues in the dataset that required correction prior to analysis.

[1] The different data collection questionnaires are briefly described in section 2.2.1 (page 12).

2.1 The Indian Case Study

The focus of this research project is on data from schools located near the Shankarpalli mandal (i.e. an Indian administrative division). For that a local partner facilitated access to the surveyed schools and the process of getting approval from the local government. The partner is the Mamidipudi Venkatarangaiya Foundation (henceforth MVF), a non-governmental organisation based in the city of Hyderabad and specialised in working with education advocacy since 1981 [MVF, 2015].

Permission to conduct the study has been sought for this research project, in cooperation with MFV. This has been granted and communicated to all selected schools headmasters through a letter dated from 30/06/2014, from the "Proceedings of the District Educational Officer – Ranga Reddy District", in which schools in respective mandals are named. Three schools in this project (namely K.school, J.school, M.school) are so-called Zilla Parishad High Schools (also known as ZPHS), meaning these are managed by the local government of the district in which they belong to. The remaining two schools (namely R.school and V.school) are private. The real names of these schools have been masked.

Following this milestone, the student parents and/or guardians (i.e. carers) have also been briefed of how this research project has been designed through letters. These documents contained brief explanations about the following topics:

- (I) communication that they have been selected to take part in the study,
- (II) offer the option for them to participate in the study or abstain from it,
- (III) that there will be no disadvantage of any kind for study participants,
- and (IV) that research outcomes may help to improve how students learn.

2.2 Data Collection

Cross-sectional data has been collected in the Indian state of Telangana, including a set of different network structures that relate to students across five schools. These range from years VIII to X and are located about an hour away drive from the city of Hyderabad. To account for the recent year's surge in private school education in this Indian state, both private and government schools have been included. The former are located in urban areas, while the latter are located in semi-urban areas.

In the dataset processed for this report there are three public schools (namely: K.school, M.school, J.school) and two private schools (namely: R.school, V.school), which are deemed representative of the area according to the Indian Unified District Information System for Education (henceforth UDISE) [NUEPA, 2015].

At the time of writing this report, various inconsistencies in the network data were being checked and fixed iteratively in collaboration with the India and Japan VAEIS project members. Further details to date (15th December 2015) are in section 2.3: "Processing The Working Dataset" (from page 14 onwards).

Ethical approval for the VAEIS project has been granted by the Waseda University Research Ethics Committee and consent forms have been collected first from schools, then from parents. Participants are free to withdraw at any time, with no need to provide a reason. The research project also had a written study protocol, for which help is offered in case collaborators have difficulties in following procedures. Other similar arrangements were in place for the partner MVF.

2.2.1 Self-Reported Student Networks

The VAEI project has been designed to collect a variety of data via the careful administration of surveys, with particular attention to the following sections:

- household: focused on gathering data about the individual student family;
- teacher: focused on gathering data about his/her individual experience;
- headmaster: focused on gathering data about the schools and teachers;
- student: focused on data about the learning environment and networks.
- academic records: focused on data about periodical tests and attendance;
- exercise sheets: focused on gathering data about an academic experiment intended to boost the overall confidence on individual academic capacity.

For this report only the network section has been used, as data from other survey sections are not yet available. Students were given a list of all names and their matriculation numbers at respective schools. Both boys and girls can be listed for one or more of the following activities: study, play and chat, class teamwork, commute, visiting after school and sharing personal problems. Each of those is a network dimension and each entry is marked as "Best Friend" or "Other Friend". For the best friends network, each entry may be further marked using a likert scale as an estimate regarding their interaction frequency per week. This ranged from: less than once, 1-3 times, at least 4 times.

The survey format, filled in on paper by each surveyed student, is available in section 5.16: "Network survey questionnaire: Best Friends. Source: VAIES project' (page 52) and section 5.15: "Network survey questionnaire: Others. Source: VAIES project' (page 51). The transcription of this data into a spreadsheet has been facilitated by MVF staff in India.

As the student networks in this case study have been surveyed according to self-reports, these are essentially perceived networks according to each individual. There is experimental evidence regarding the level of information accuracy when people are asked to self-report their own past. Generally accuracy increases with interaction patterns containing dyads (i.e. A→B) embedded within triads (e.g. A→B, then B→C and then C→A interactions)[2]. Social interactions that are structured or involving rich experiences [H R Bernard et al., 1984, Freeman et al., 1987, Brashears and Quintane, 2015]. In other words, individuals tend to not remember well interactions lacking clear structures –even when accounting for frequency. This is in line with the sociology-driven theoretical underpinnings of social network analysis, [Granovetter, 1982, Wasserman and Faust, 1994, Burt, 2004, Robins et al., 2007, Snijders et al., 2010, Lusher et al., 2012]:

- which pose that social networks are locally emergent (i.e. ties between individuals are formed by social process(es) that are mostly locally determined);
- social networks self-organise (i.e. dependencies are between individual local links and these individual attributes, not just –or at all– their aggregation);
- patterns within social networks require at least 2 individuals engaged in at least 1 social interaction pattern, which is evidence of a formation process.

Therefore social networks depends on understanding patterns of local linkages (e.g. types of dyads and triads) in terms of their relevance to collective social structures. An observed social network is then expected to be generated by the one or more local process that result from the self-organising relationships between individuals. This is a stochastic process (i.e. not fully predictable) but still can be studied to describe and infer –with some degree of error– the network structure. In this sense there are a number of dyad-level dependencies (e.g.: reciprocity, transitivity, homophily, paths, circuits, etc.) that are essential to include in social network analysis to account for social structure and interaction[3].

[2] Transitivity is discussed further on page 22 and censuses on pages 44, 45, 46 and 54.
[3] These implications are discussed further in section 3: "Research Findings' (page 21).

2.3 Processing The Working Dataset

In order to uniquely identify each student across the dataset files, it has been necessary to create an identification key for each of them. The only way this could be achieved was to use the following combination per student entry: school, class, section and matriculation. This extra variable has been named "global_ID" and has been created in all files after inconsistencies in data entry have been corrected.

Hence the dataset received and processed for this report consists of 3 files, namely:

- "cover.prn" with 17 variables serving as an index for individual student data. This include: serial number, household, number, name, gender, age, village, home, religion, class, section, teacher, supervisor, survey date month hour and minute. The file has 1072 data rows and missing values included: village (18), home (12), section (84), number (12) and household (87).

  Further 82 other missing entries for section have been replaced to contain "a (telugu)" for the school M.school – which could only contain such section according to MVF. After removing non-retrievable entries, the dataset reduced to 1060 rows. The remaining missing data reduced to these entries: village (16) and home (11)[4]. The table below show the school composition.

Table 2.1: Gender: sections, schools (private: R.school, V.school) Source: author

		section	a (telugu)	b (english)	c	Total
1	school	gender				
2	J.school	female	52	36	0	88
3		male	24	42	0	66
4	K.school	female	86	64	0	150
5		male	76	60	0	136
6	M.school	female	49	0	0	49
7		male	33	0	0	33
8	R.school	female	52	48	12	112
9		male	56	56	16	128
10	V.school	female	40	56	30	126
11		male	67	62	43	172
12	Total	female	279	204	42	525
13		male	256	220	59	535

[4] These variables have not been used, so could feature in future work.

- "net.prn" is the file containing entries regarding the best friends survey, with 10720 row entries, 19 variable codes, namely: serial, school, class, section id, c, s, a, n, sy, sf, py, pf, ty, cy, vy, vf, hy. These codes correspond to those shown the verbatim copy of the administered survey[5]. This file is coded in a somewhat unusual way for network analysis, as data is often stored either an adjacency matrix (i.e. 0 when a link is absent and 1 otherwise) or an edge-list (i.e. a chain of links without missing identification values). The file "net.prn" is akin to an edge-list that requires extensive consistency checks: within itself and in relation to the index of surveyed students, which is available in the previously discussed "cover.prn" file.

  Consistency checks are needed in this case as the unique identification within the network dataset is a combination of values. Hence erroneous indices required fixing within "net.prn" before retrievals could be attempted based on "cover.prn". These peculiarities are depicted in the table below. In network analysis it is needed to know who is A and who is B (i.e. one person is linked to another: A→B, as one would not socially link A→A).

Table 2.2: Missing data in the file "net.prn" (best friends). Source: author.

variable	Missing data
section	840
id	0 entries, yet 10600 marked as 0
c	3839
s	4314, yet 4 marked as e
a	3843 missing, yet 6876 marked as 0
n	3532
sy	5928
sf	6453
py	6569
pf	7254
ty	7127
cy	7417
vy	7926
vf	8350
hy	7309

[5] Please refer to it on section 5.16: "Network survey questionnaire: Best Friends. Source: VAIES project" (page 52).

The same correction applied in “cover.prn” has been applied in “net.prn”. That is: if the variable school is equal to M.school and variable section, or variable s, is equal to c, the value is replaced with “a (telugu)”. When the variable s is equal to e, the value has been replaced to “b (english)”. Once again these instructions have been followed according to advice from MVF.

To facilitate working with the dataset, variables have been renamed according to the table below and then converted to binary values (i.e. if there is a missing value, entries are replacement with 0, or “no”, for each variable).

Table 2.3: Original and new variables in “net.prn” (best friends). Source: author.

	original	renamed
1	sy	net_study
2	sf	net_study_freq
3	py	net_play_chat
4	pf	net_play_chat_freq
5	ty	net_class_team
6	cy	net_commute
7	vy	net_visit
8	vf	net_visit_freq
9	hy	net_personal_share

Upon fixing the missing values according to consistency checks within “net.prn” and “cover.prn”, the final table of non-retrievable values is shown below.

Table 2.4: Fixed missing data in “net.prn” (best friends). Source: author.

variable	Remaining missing data
c	252
s	691
n	2
net_study	2319
net_study_freq	2726
net_play_chat	3077
net_play_chat_freq	3516
net_class_team	3571
net_commute	3965
net_visit	4541
net_visit_freq	4541
net_personal_share	3726

Having completed the consistency check within the file, global_IDs have been created for each survey respondent and each named person in the survey. This meant the creation of two new variables in “net.prn”, namely:

- global_egoID, a key per unique combination of values: school, class, section, id variables. This is used to identify each survey respondent.

- global_alterID, a key per unique combination of values: school, c, s, a variables. This is used to identify each person named in a survey.

In social network analysis terms, global_egoID is equivalent to A, while global_alterID is equivalent to B in links such as these: A→B and B→A. At this point then one can consider which network structure to work with. In total 228 global_alterIDs still contained missing values and had to be removed from “net.prn” [6]. The total entries in “net.prn” reduced to 1150 links present across activities, with 311 unique egos and 603 unique alters.

These may be either an undirected network (i.e. A→B = B→A) or a directed network, also known as a di-graph (i.e. A→B $\neq$ A→B). The implications for choosing one type or the other are significant as a non-directed network essentially means the diagonal of the network adjacency matrix [7] will mirror the structure (i.e. the upper diagonal part will be the same as the lower part). This will lead to faster calculations (as only half of the matrix has to be processed) and will change network statistics, given that it is not needed to account for the directionality of every present link.

[6] Some of these may be retrievable as those in section 2.3.1: “Fixing “other_network.prn” with “cover.prn”” (page 19), yet is future work as the MVF partner would need to confirm this.
[7] i.e. the matrix representation (1 if link is present, 0 otherwise) of a network edge list.

- "other_network.prn" is the file containing entries for friends other than one's best, so it should not contain people listed in "net.prn". There are 32100 original entries, with missing data as depicted in the table below. There are 11 variables: serial, school, class, section id, v, c, s, a, n, rank.

Table 2.5: Missing data in "other_net.prn" (other friends). Source: author.

variable	Missing data
id	0, yet 360 are marked as 0
c	18978
s	19941, yet 1067 marked as c
a	19351
n	18630

The same aforementioned correction regarding the school M.school has been done, resulting in the replacement of 2490 missing values in section with the following value: "a (telugu)", according to the MVF partner.

Table 2.6: Fixed missing data in "other_net.prn" (other friends). Source: author.

variable	Remaining Missing data
c	18766
s	19686
a	19112
n	18425

974 unique keys are created as global_egoIDs using the combination of: school, class, section, id. And 2116 unique keys have been created as global_alterIDs using the combination of: school, c, s, a. Then the variable v has been renamed to network to facilitate the analysis. As a consequence of missing data, 18433 global_alterIDs are non-retrievable (i.e. lack 3 identifiers in each key, meaning that the only identifier left was the school a student goes to).

2.3.1 Fixing “other_network.prn” with “cover.prn”

In order to attempt resolving at least part of the missing data in both network files, a secondary key has been created based on the following variables from “cover.prn”: school, class, name. The same key pattern has been created based on the following c from “other_network.prn”: school, c, n. This key has been created as auxSchoolClassNameID in all 3 files.

Upon removing the non-retrievable keys, “other_network.prn” remained with 13307 entries. From these 86 have 2 missing identifiers (s, a) in its keys, which may be recoverable from “cover.prn” using the secondary key. 67 of these did not match, so are removed as identifiers are still missing. The remaining 22 entries have been fully matched, so their global_alterID keys are updated accordingly. After this, one further error in the original dataset file is found and deleted: 1 entry where global_egoID = global_alterID. In total then 22 entries have been merged back into the “other_network.prn” data, with 10 unique global_egoIDs and other 14 unique global_alterIDs.

In total there were still 526 cases where a key contains the a identifier and 69 of these also missing the c. 84 have been fully retrieved, which required updates of respective global_alterIDs required updating in “other_network.prn”. Yet then again one more case of global_egoID = global_alterID appears and is dropped. The updated dataset then grows from 11458 to 11562. Then 33 entries in the subset containing other friends only are deleted due to self-references, resulting in a network size of 9994. Further 4 entries are deleted due to self-references in the subset containing only confidants.

2.3.2 Cross-Activity Sub-set of Best Friends

Despite the aforementioned limitations, sub-setting the network data without missing values across activities does allow the creation and gathering of insights from their sub-structures. The table below is a summary of the network composition by activity among best friends.[8] This is the dataset that has been analysed for this report as an undirected network across activities (i.e. A→B suffice).[9]

Table 2.7: Undirected network size by activity (best friends). Source: author.

	study	play / chat	class team	commute	visit	size
1	yes	yes	yes	yes	yes	1157
3	yes					768
4	yes	yes	yes			416
5		yes				323
6	yes	yes	yes	yes		322
7	yes	yes				318
8	yes		yes			220
9				yes		195
10			yes			170
11	yes	yes		yes	yes	136
12	yes	yes		yes		117
13	yes	yes	yes		yes	114
14					yes	106
15	yes			yes		90
16	yes		yes	yes		89
17	yes		yes	yes	yes	60
18		yes		yes		58
19		yes	yes	yes	yes	53
20	yes	yes			yes	52
21		yes	yes			51
22	yes			yes	yes	44
23		yes	yes	yes		34
24			yes	yes		34
25		yes		yes	yes	33
26	yes		yes		yes	31
27		yes			yes	30
28			yes	yes	yes	27
29		yes	yes		yes	23
30				yes	yes	22
31	yes				yes	21
32			yes		yes	9

[8] A similar but more complete table, per school, is available in section 5.29 (page 75).
[9] A more detailed explanation regarding this choice can be found on page 21.

3. Research Findings

In this case study, only the personal sharing (confidant) network can be considered indisputably directed (i.e. A chooses to share a private problem with B, but B can choose to not share with A). All other network dimensions (i.e. play and chat, class teamwork, commute, visiting after school and sharing personal problems) can be interpreted as non-directed (i.e. if A reports B does play together, it is reasonable to assume this suffice and hence there is no need for B to confirm).

The dataset has thus been weakly symmetrised [Butts, 2008][1], which is a conservative way to work with the task-oriented network dimension. The task in question can be interpreted as non-directional, so a single perceived link is sufficient for the link. Thus the best and other friends network is hereby analysed as an undirected activity network (i.e. A→B suffice, meaning focus is on tasks).

This report only contains data from the network section of the administered questionnaires, as previously discussed[2]. Each of following section is dedicated to a subset of friends, namely: best, others and confidants. To facilitate, all presented tables are also available in a combined form in section 5.4 (page 58).

[1] An alternative that would likely change the network structure considerably is a stricter symmetrisation, such as that a link should only exist if both ways are reported (i.e. the case of A→B and B→A), or that the upper/lower half of the adjacency matrix is be copied.

[2] Please refer to section 2.2: "Data Collection" (from page 11) and also to section 2.3: "Processing The Working Dataset' (from page 14).

3.1 Fundamental Analytical Network Concepts

In order to understand the analyses carried out in each network subset (i.e. best, other and confidant friends), this subsection is dedicated to introduce and briefly discuss concepts that have been used to calculate network properties. It follows:

- **Transitivity**. This property considers patterns of triples in the following form: every time A→B and B→C, then A→C may hold [Wasserman and Faust, 1994]. The adage "friends of my friends are my friends" does encapsulate the concept of transitivity, as the probability of nodes being connected within the same set [Snijders et al., 2006]. This is a complementary concept to clustering coefficient[3] as the latter is a measure of how much nodes tend to be clustered in a network and experience closure [Chen and Yang, 2010]. These concepts can be useful to understand network properties with regards to its overall density and / or sparsity. Hence, higher levels of transitivity tend to increase the **density** (i.e. ratio in terms of actual number of links and the number of possible links in a network) and the average number of links. Transitivity and density can thus help with the identification of smaller, distinct groups in a large network, in the form of clusters (i.e. given the level of graph clusterability) or cliques (i.e. different definitions of subgroups) [Davis and Leinhardt, 1967]. Node triads, particularly transitive ones[4], encourage interactions within subgroups and partition sets. Directed graphs have 16 types of triads and undirected four.
  - **Vacuously transitive and intransitive**: If one of the two conditions stated above for transitivity is not met, then the triple is classified as vacuously transitive. The antonym of a transitive triad is intransitive[5].

[3] Both in local (single-node) and global (overall) levels of network embeddedness.
[4] These can be the simplest structure suggesting hierarchy [Hanneman and M., 2005].
[5] These three concepts are further discussed on page 43 and also illustrated in figure 3.11.

- **Component**. This is also known as a connected graph, containing a path between every possible node (i.e. a component without disconnected nodes). Hence every element in it can be reached from any starting point in the network. When a graph is disconnected, it is possible to partition the graph into several sub-graphs that are connected, thus allowing to count the number of components and their sizes [Chen and Yang, 2010]. With this in mind, the largest component can be interpreted as a measure of graph reachability, with implications regarding possible shortest paths, diffusion dynamics and likelihood of network attachments (e.g. group membership and growth).

  - The **largest component**, which is also referred to in the literature as giant component, of a network usually contains special properties due to its unique structure. For this reason, often researchers choose to sub-set it altogether and analyse it independently from other nodes in a network [Hanneman and M., 2005]. This could be a useful comparison yardstick in the context of this project, as often there are disconnected school nodes. It is important to note that this measure is particularly sensitive to symmetrisation (i.e. graph simplification) procedures.[6]

- **Path**. This is the number of unrepeated links in a graph, or network, sequence [Chen and Yang, 2010]. The path length between any two nodes can thus be measured and the average path can serve as a complementary measure of reachability (as it accounts for distances between all node pairs).
  - **Diameter.** This is the furthest possible path in the graph or network.[7]
  - **Reciprocity**. If the link A→B exists, does B→A also? This explicit differentiation is necessary in networks where A may contact B, but not vice-versa (e.g. advice-giving, lending money) [Granovetter, 1982].

[6] That is: weak, strong or diagonal copies, as discussed on footnote of page 21.
[7] If the network is unconnected, then the diameter refers to the largest geodesic, or largest shortest path, in the largest component of the network.

- **Assortativity**. This coefficient indicates the level of network structure similarity, using a degree correlation of adjacent vertices, thus ranging from positive 1 (maximum similarity, i.e. segregated) to negative 1 (least similar) [Bojanowski and Corten, 2014]. This can be helpful to test the assortativity level among best-achieving students with regards to each surveyed network dimension –similarly as in [Vu et al., 2015]. That is, this concept allows to probe into the extent nodes of certain types are linked with other similar nodes; both through structurally-driven and categorically-driven selective linking [Kolaczyk and Csárdi, 2014]. In modelling networks, often this coefficient is useful to understand the overall influence in the tested model. That is because preferences of nodes may depend on their degrees; both incoming (in-degree) and outgoing (out-degree) properties as such:

  (a) out ego-out alter; (b) out ego-in alter;

  (c) in ego-out alter and (d) in ego-in alter [Snijders and Koskinen, 2011].

  These are found in [Ripley et al., 2013] and represent approximate representations of mixing patterns (i.e. network tendencies to link up with what type of individual). Some seek those who are similar: assortative mixing (i.e. homophily) as typically seen in collaborative networks; while others seek who are different: dissortative mixing (i.e. heterophily), for example websites that may be about the same topic but using unrelated references [Dorogovtsev and Mendes, 2004]. Therefore one could argue that certain dyadic mixing patterns, but particularly at the triad level, will contribute towards the consolidation of certain network patterns [Newman, 2003].

  Dissortative networks often present high-degree nodes and so are susceptible to bottlenecks through such nodes and other vulnerabilities as a consequence of the link concentration [Kossinets, 2003, Goodreau et al., 2009].

Another important set of concepts in network analysis is that of centrality measures. There are a number of these, each dedicated to highlight a specific sociological aspect in a network, such as: prestige, influence, bridging roles [Newman, 2010]. Perhaps the most commonly used measures of centrality are the following:

- **Closeness.** This centrality measure takes into account the immediate links of a node, thus ignoring every indirect link (i.e. through other nodes) in a network [Freeman, 1979]. For this reason it tends to highlight central nodes in neighborhoods (local or global) [Hanneman and M., 2005]. It can be interpreted as the average geodesic distance from a node to all other nodes in the graph. This means of a way of identifying those that are overall best connected [Faust, 1997]. In short it can be a measure of how close a node is to others in the network. That is, one is central if can interact quickly. This notion is inspired in early operations research and psychology works such as [Hakimi, 1964, Sabidussi, 1966], where the concept of "minimum steps" was used to discuss what is now generally referred to in network analysis as geodesic distances (i.e. how further one is in relation to others?).

  In directed graphs, this measure can be more meaningful if separated into: in-closeness and out-closeness, to account for differences of incoming and outgoing links. Hence the independence, or inter-dependence, of nodes may become particularly important in disconnected networks (i.e. where there are no links between all nodes) [Zhang, 2010, Wasserman and Faust, 1994]. Therefore it is useful to understand both the average closeness centrality, its standard deviation and the per-node result of such closeness centrality. In this way one can better understand the range of closeness centralities.

  The results of closeness centrality, for both undirected and directed networks per school, are provided in tables 3.10, 3.12 and 5.26 (pages 32, 35 and 64).

- **Betweenness.** This centrality measure gives a sense of strategic positioning within a network, as it accounts for the number of linked pairs that would otherwise not be linked without bridging nodes [Anthonisse, 1971, Freeman, 1979]. That is, the more shortest paths through a node, the higher this centrality measure will be for that node. This requires the calculation of every shortest path between every pair in the network. This can be informally interpreted as traffic passing through nodes, helping to depict indirect influence in networks due to highlighting intermediaries that mediate flows between others. Therefore this measure provides a finer-grained understanding of the extend in which nodes may exert control over other nodes according to their position in a network [Everett and Borgatti, 1999].

  Hence this measure is not about who is central in terms of directly incoming or outgoing links, but in terms of who provides indirect links (bridges) between different groups or network structures. High betweenness then is only possible in networks that have relatively low density levels, where one can best probe and identify nodes functioning as gatekeepers or brokers. Nodes with high betweenness centrality are bottlenecks, as link non-adjacent nodes indirectly through their intermediary positions. This allows them to function as facilitators or interruptions in the flow of information. This thus highlights the potential roles of inter-personal influence and the level of heterogeneity in network mixing patterns [Wasserman and Faust, 1994].

  A variation of betweenness centrality is that of flow, where all links between nodes in a given shortest path are required. This development moves away from the original definition, aforementioned at the top of this page, as then the purpose is to focus on the usefulness of calculating flow paths that depend on direction (i.e. in di-graphs) [Hanneman and M., 2005, Borgatti, 2005].

- **Eigenvector**. This centrality measure is based on the notion of that a node is central if is also connected to others, also relatively central nodes. This means that identifying central nodes according to the eigenvector is essentially an exercise that also takes into account the centrality of its neighbours [Bonacich, 1972]. In other words, the relevance of one's direct contacts is what is being valued most. There are different social networks centralities measures that are similar in terms of also taking into account one's neighbourhood (i.e. walk-based or degree-weighted score approaches) [Smith and Moody, 2013], such as [Katz, 1966] centrality, Hubbell centrality [Hubbell, 1965], Bonacich power[8] (i.e. the power of a node being recursively defined by the sum of the power of its alters[9]) [Csardi and Nepusz, 2006]. For a non-social variation of this concept, there is the well-known Page Rank algorithm [Page et al., 1999], which has been further developed at Google.com to classify the relevance of websites in relation to others for online search using keywords. Thus this concept is useful to understand network node prestige / status, overall levels of relevance and cooperation [Butts, 2007, Rodan, 2011][10].

- **Degree**. This centrality measure is in direct contrast with the previous definition, as it only takes into account the number of connections a node has –with no regard about the centrality of its neighbours [Borgatti, 2005]. In this sense then degree centrality is concerned only with the immediate risk or potential related to a particular node, while the eigenvector –and its variants– is each concerned with both direct and indirect node effects.

[8] This is also referred to as alpha centrality or beta centrality; being interpreted as allowing access to network flows indirectly (i.e. without directly maintain them). [Smith et al., 2014].

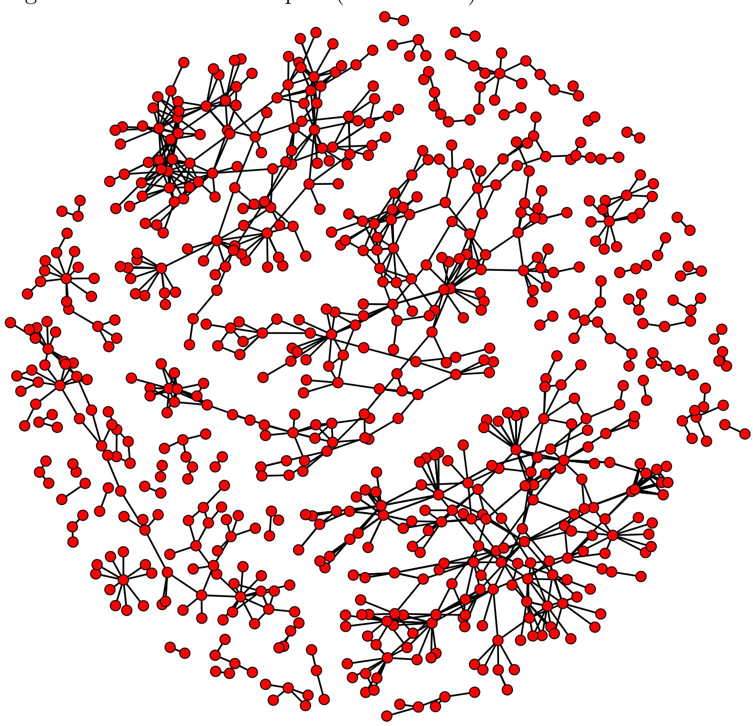
[9] In other words, the more powerful one's direct neighbours are, the more power one has.

[10] For example: (1) [Banerjee et al., 2013] hypothesised that villages where first contacted people have higher eigenvector centrality, there should be a better spread of information; and (2) [Fowler et al., 2008] hypothesised about the spread of happiness is influenced by the eigenvector centrality as a measure of importance to a person's local network –rather than the whole.

3.2 Best Friends across Activities

The overall network structure of the subset containing students nominated as best friends, across activities, is shown in the figure below. One can visualise a number of groups, varying in size, that remain disconnected from others. Some of these are due to students being in different schools, while others are within the same school. These differences are discussed next, with focus on each school.

Figure 3.1: Plain network plot (best friends) Source: The author.



The best friends subset has been analysed as an undirected graph, so one should bear this in mind when interpreting the presented results. If a stricter assumption had been taken (i.e. both A→B and B→A must appear in the survey data), then the appropriate approach would be the one discussed as future work (page 49).

The composition o best friends networks, for each school, is shown below. [11]

Table 3.8: Network size per school, section, class (best friends). Source: author.

	school	section	class	8	9	10	Total
1							
2	J.school	a (telugu)		195	50	285	530
3		b (english)		445	265	75	785
4		c		0	0	0	0
5	K.school	a (telugu)		385	470	270	1125
6		b (english)		185	620	170	975
7		c		0	0	0	0
8	M.school	a (telugu)		0	235	0	235
9		b (english)		0	0	0	0
10		c		0	0	0	0
11	R.school	a (telugu)		105	125	145	375
12		b (english)		110	330	290	730
13		c		70	0	0	70
14	V.school	a (telugu)		195	160	140	495
15		b (english)		155	40	110	305
16		c		110	15	0	125

Most schools contain students in every class (8, 9, 10), section (english, telugu). Section “c” is only present in the private institutions, R.school and V.school. M.school only contain students in class 9, section telugu –which is a consistent limitation, appearing also in the subset of other friends and confidants[12].

Given the typical importance of the largest component in social networks, the analysis in this report has been done by amalgamating sections and classes per school. This is a way to facilitate understanding the overall structure per school in relation to the most connected set of nodes. Research findings from the aforementioned procedure to the survey data are presented in the table next [13].

[11] For the full subset composition per activity, refer to the table in section 5.27 (page 67).
[12] See memberships in section 5.5: ”Summary of Best, Others and Confidants” (page 59).
[13] A future work could analyse the school network structures per class and section.

The table below depicts the overall structural properties for the network of best friends in each school, across activities. Reciprocities are all equal to one due to the network being undirected (U). Otherwise, if directed, it is shortened to (D)[14].

One can observe that most LC encompass most of the total number of students, with the exception of the last entry –which is rather disconnected in comparison to other schools. This suggests that most schools contain most of its surveyed cohort accessible across the activities –albeit with a somewhat moderate variation of average path length (from 3 to 7). Most schools have an average of 5 students in the path for reaching any other student in the network across activities.

Table 3.9: Summary statistics (best friends). Source: author.

	#S	#L	Trans.	Path	#C.	LC.	D.	Den.	A.	RC.
J. Best Friends (U)	148	244	0.262	5.143	8	131	13	0.022	-0.050	1
K. Best Friends (U)	188	294	0.190	5.002	4	180	10	0.017	-0.370	1
M. Best Friends (U)	25	42	0.369	3.167	1	25	6	0.140	-0.488	1
R. Best Friends (U)	174	225	0.215	7.017	15	132	19	0.015	0.028	1
V. Best Friends (U)	191	179	0.071	5.415	28	63	15	0.010	-0.332	1

Legend:
Path = average path, #C = number of components, D. = diameter
LC = largest component, Den. = density, RC. = reciprocity ratio
#S = number of students, #L = number of links, A. = assortativity
Trans. = transitivity

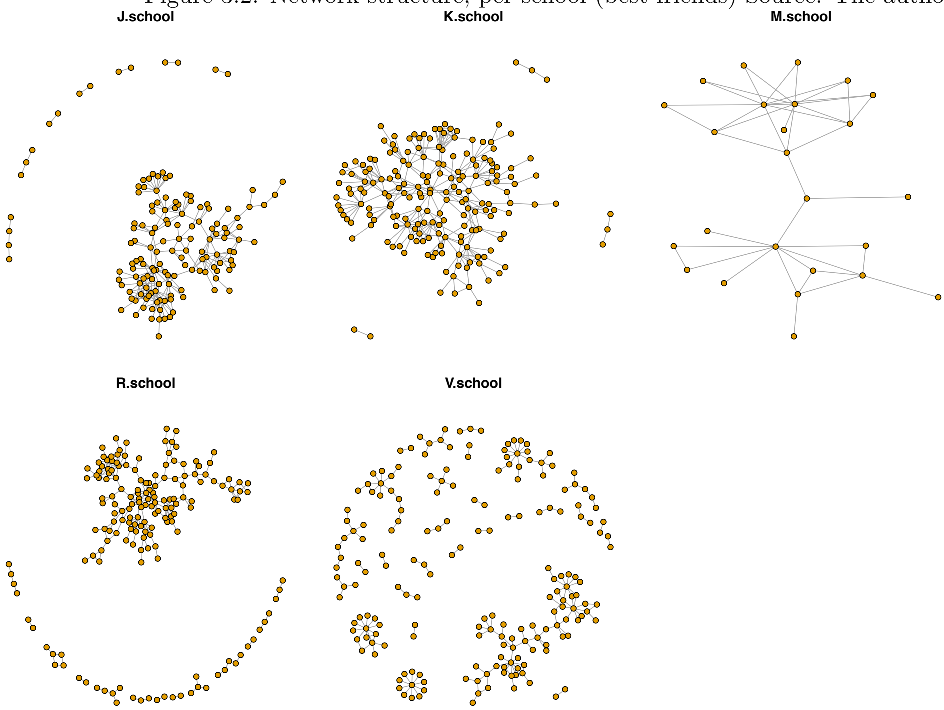
The transitivity values are positive but not large, which denote a mild tendency for hierarchies within the surveyed schools. This is further suggested by the wide variation in the number of components (from 1 to 28) and overall mild negative assortativity values (i.e. tendency for less network homophily). These results are similar for the other network subsets (i.e. other friends and confidants).

[14] This notation will be the same used in all others appropriate tables found in this document.

The [Kamada and Kawai, 1989] algorithm has been chosen to plot these networks as it is undirected and so it is helpful to understand its graph-theoretic distances. This is done by using a theoretical force-based plotting approach to rank and then map empirically faraway nodes to be displayed with greater distances. This facilitates visualising core-periphery structures in networks [Kobourov, 2013] as:

- J.school and R.school present the classical core-satellites components [15] ;
- K.school is somewhat similar to the aforementioned schools, but with fewer isolated components and typically these contain cores with higher density;
- V.school contains various components (mostly small structures) while M.school has one component and a bridge between two visually identifiable groups.

Figure 3.2: Network structure, per school (best friends) Source: The author.



[15] A sensitivity-based layout could be used, where links are simplified into a skeleton network with the original structural properties [Crnovrsanin et al., 2014] (see section 4.1, page 50).

The table below depicts three centrality measures for the non-directed (U) network of best friends, per school. It is interesting to notice that betweenness and closeness centralities do not considerably across the subsets, despite some occasional relatively large deviations. On the other hand, the eigenvector measure contain some very large variations, which may be a result of the network having relatively short average paths throughout (see table 3.9 on page 30) but few nodes are found to be positioned strategically enough to rank very high in this measure.

Table 3.10: Summary of network centralities (best friends). Source: author.

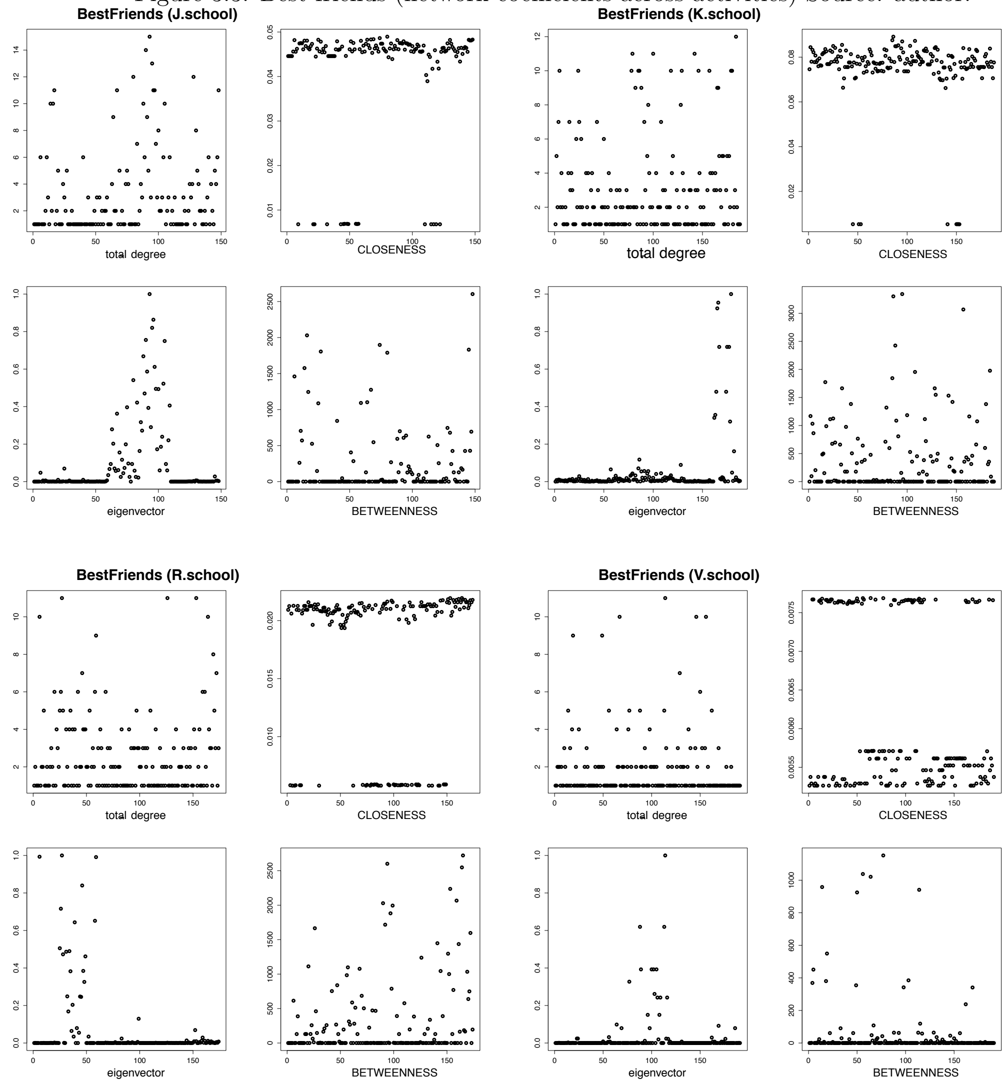
	Betweenness	Closeness	Eigenvector
J. Best Friends (U)	0.042 (0.013), 0.046	0.099 (0.203), 0.001	238.736 (484.939), 0
K. Best Friends (U)	0.075 (0.015), 0.078	0.048 (0.16), 0.006	343.053 (616.465), 0
M. Best Friends (U)	0.324 (0.054), 0.304	0.271 (0.315), 0.133	26 (51.959), 0
R. Best Friends (U)	0.017 (0.007), 0.021	0.064 (0.189), 0	300.69 (574.482), 1
V. Best Friends (U)	0.006 (0.001), 0.006	0.032 (0.118), 0	56.356 (190.943), 0

Format: Average centrality (standard deviation), median

R.school, J.school and K.school, despite having a rather clear visual suggestion of containing core-periphery structures (please refer to figure 3.2 on page 31), are structured in a way which ends up providing a similar picture overall with regards to the centrality measures of closeness and betweenness. The other networks are also similar in this regard [16]. However the shape of these structures does not seem to visually suggest, as clearly, whether core-periphery could be a reasonable assumption for specifying testable models taking into account this feature for explaining network formation and / or maintenance of its linkages.

[16] With the exception of M.school, which due to its rather small size may require another approach in order to better understand its structure and relationship to academic achievements.

Figure 3.3: Best friends (network coefficients across activities) Source: author.

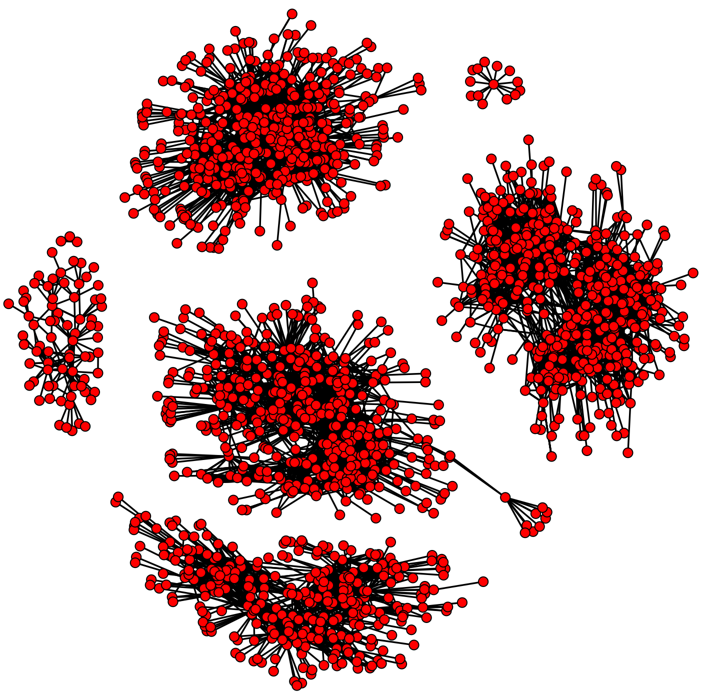


Networks have similar properties regarding: degrees and betweenness (most nodes with low coefficients and few relatively large), closeness (most nodes with either moderate or low coefficients) and eigenvector (most nodes with low coefficients). M.school may require a different approach for analysis due to the rather small sample: 25 (see figure 5.21, page 56, depicting properties as in the figure above).

3.3 Other Friends across Activities

The figure below depicts the overall structure of the network consisting of other friends, across activities, in the VAEIS survey. There are six components (i.e. clusters that are disconnected from other nodes). This suggests that in one school there is a group of students that either is segregated from others or who have not participated. This check is ought to be done by the VAEIS partner, MFV.

Figure 3.4: Plain network plot (other friends) Source: author.



This network has also been analysed as undirected using the survey data as given. An alternative symmetrisation would be to check whether those appearing in the chosen version of the symmetrised best friends network[17] do not appear in the subset of other friends. Note that this could be done across or per activity.

[17] That is: weak, strong or diagonal copies, as discussed on the footnote of page 21.

It is surprising that the average path and density for other friends are even more similar throughout than for best friends. This may suggest that symmetrising these in stricter ways would change them considerably, as then the largest component would likely not encompass as many nodes and the diameter may increase.

Table 3.11: Summary statistics (other friends). Source: author.

	#S	#L	Trans.	Path	#C.	LC.	D.	Den.	A.	RC.
J. Others (U)	311	1060	0.238	3.507	2	308	7	0.022	-0.103	1
K. Others (U)	445	1927	0.221	3.372	2	442	7	0.020	-0.025	1
M. Others (U)	97	135	0.104	3.288	5	70	8	0.029	-0.328	1
R. Others (U)	422	1914	0.258	3.629	1	422	9	0.022	0.078	1
V. Others (U)	431	2073	0.219	3.454	1	431	7	0.022	0.052	1

Legend:
Path = average path, #C = number of components
LC = largest component, D. = diameter, RC. = reciprocity ratio
#S = number of students, #L = number of links, A. = assortativity
Trans. = transitivity

The table below depicts three centrality measures for the non-directed (U) network of other friends, per school. Once again, as in the best friends sub-set, there are general similarities with regards to the betweenness and closeness centralities. Eigenvector in this case does not vary as much as in best friends, which is expected due to the greater diversity of social interaction in contexts involving different types of friendships –as perceived by each surveyed individual student[18].

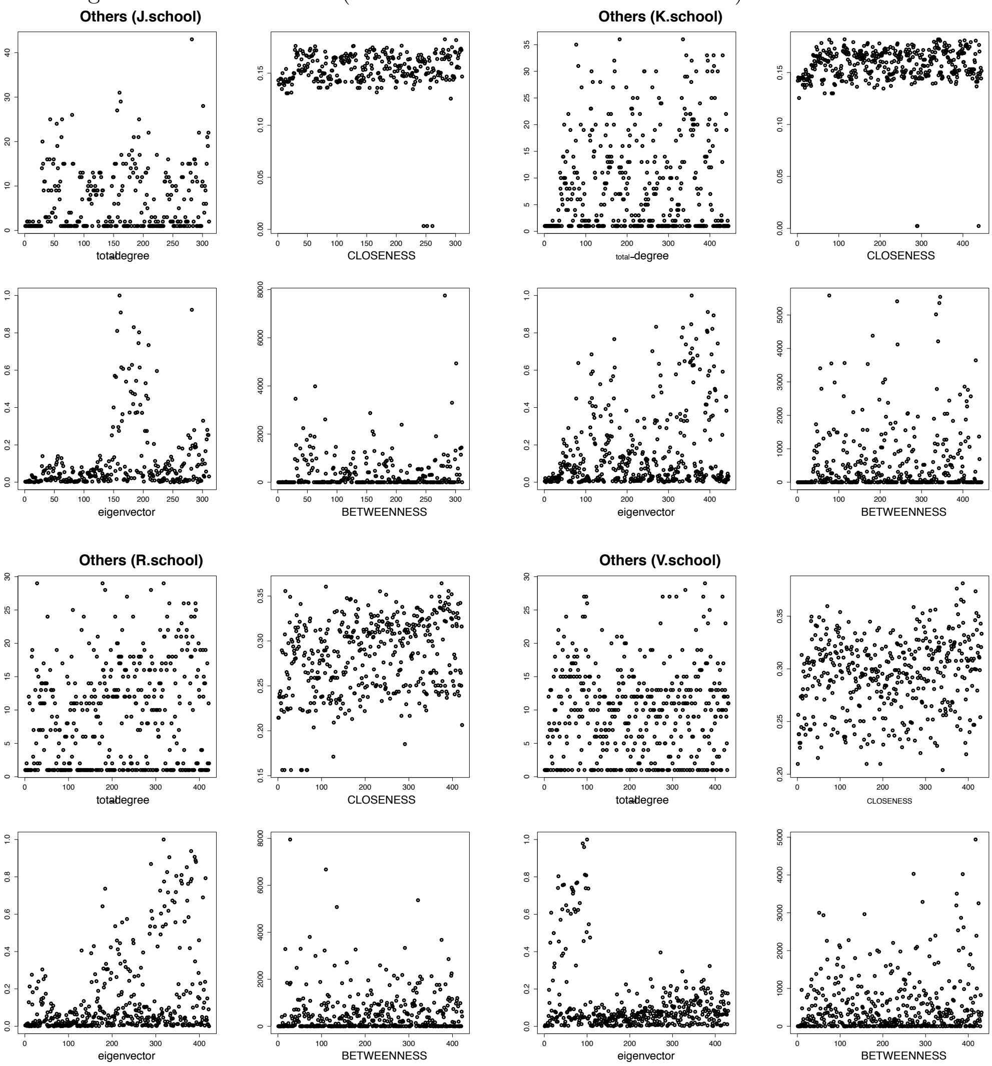
Table 3.12: Summary of network centralities (other friends). Source: author.

	Betweenness	Closeness	Eigenvector
J. Others (U)	0.154 (0.019), 0.154	0.111 (0.182), 0.044	381.148 (782.894), 13.375
K. Others (U)	0.157 (0.018), 0.16	0.16 (0.212), 0.067	519.622 (951.318), 50.076
M. Others (U)	0.027 (0.01), 0.033	0.141 (0.22), 0.06	59.876 (134.919), 0
R. Others (U)	0.283 (0.042), 0.287	0.145 (0.22), 0.047	553.436 (912.947), 197.987
V. Others (U)	0.294 (0.035), 0.298	0.121 (0.182), 0.062	527.603 (740.479), 266.685

Format: Average centrality (standard deviation), median

[18] It remains future work checking whether there are differences in centralities when this set is symmetrised differently (i.e. weak, strong or diagonal copies, as in footnote of page 21).

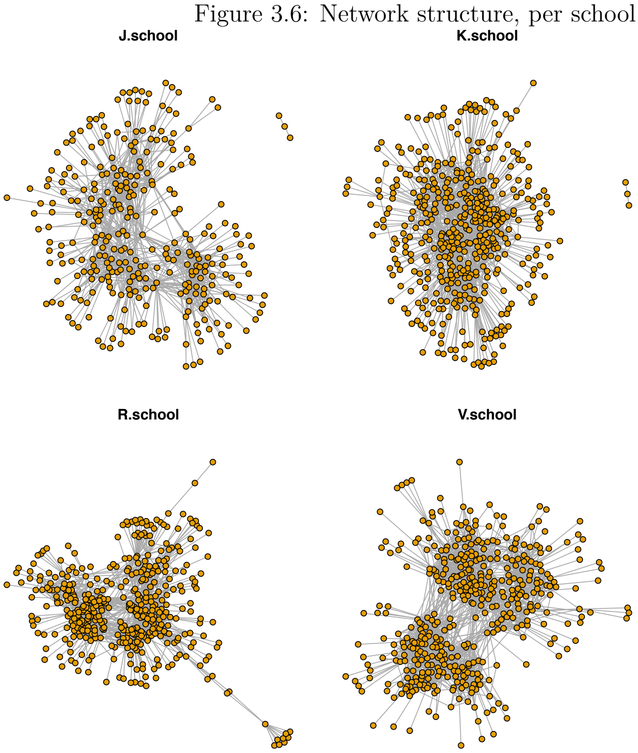
Figure 3.5: Other friends (network coefficients across activities) Source: author.



The figure above depicts the wide variation in network configuration of coefficients for four schools. Once again, the network for M.school is rather small (97) and this would require another approach, specific to it. With the following exceptions: (a) closeness in K. and J. schools is mostly concentrated within a small range; (b) betweenness coefficient values are often low, apart from K. and V. school.

The [Fruchterman and Reingold, 1991] algorithm has been chosen to plot this network for similar reasons as the approach for best friends (figure 3.2, page 31). This is also a force-directed approach, thus belonging to the group of spring-embedder approaches, but with the addition of attraction between the closest nodes –which tend to generate clearer overall images with greater graph density[19]. In this case the image below illustrates the overall pattern within the schools networks of a more densely connected core with various peripheral nodes[20]. Out of the 881 unique egos and 1670 unique alters, only 36 egos identifications are not present as alters and only 2 of those are not present in the "cover.prn" index.

Figure 3.6: Network structure, per school (other friends) Source: author.



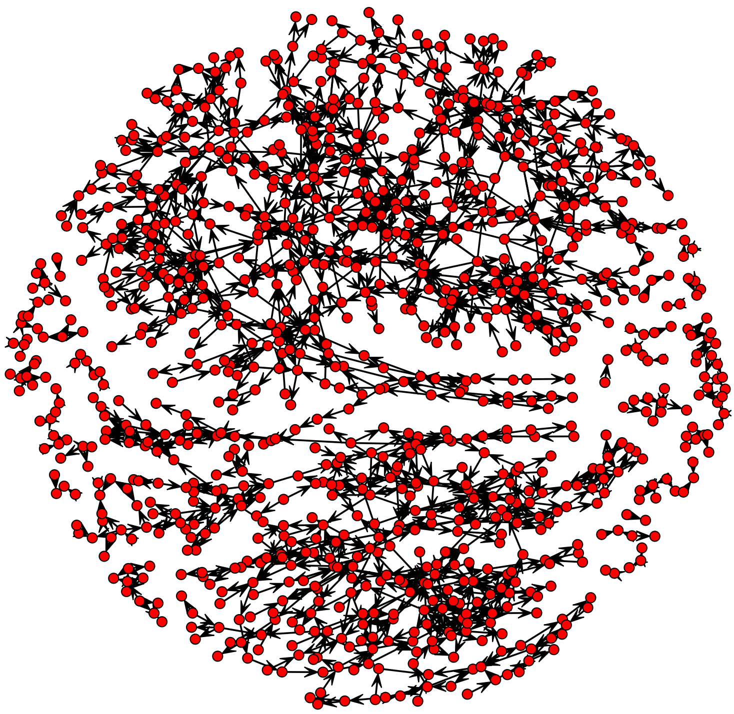
[19] In a simplified analogy, this is akin to each node having a magnetic repulsion from others yet links between them behave as springs that pull them together in the system [Chang, 2012].

[20] An alternative to avoid potential local minima is a 2-stage process: break tied values randomly from a geodesic distance matrix and then apply the springs [Wilkinson et al., 2005].

3.4 Confidant Networks

The plot below depicts the overall structure of the confidant network across schools. One can visualise various rather long directed chains of connected individuals and also various small groups, including too the cases of exclusive dyads (that is, node A is linked to node B but to nobody else in the whole network).

Figure 3.7: Plain network plot (confidant friends) Source: author.



The next pages are dedicated to further understand these structures at the school level. In doing so, particularities of certain schools are revealed and common features across them are also discussed. This is done by taking into account both a triad census per network and their respective properties such as centralities.

Table 3.13: Summary statistics (confidants). Source: author.

	#S	#L	Trans.	Path	#C.	LC.	D.	Den.	A.	RC.
J. Confidants (D)	148	167	0.123	2.284	15	107	8	0.008	0.312	0.025
K. Confidants (D)	261	351	0.089	5.639	13	234	14	0.005	0.094	0.012
M. Confidants (D)	33	28	0.176	1.316	7	10	3	0.027	0.048	0
R. Confidants (D)	277	404	0.109	7.487	10	250	25	0.005	0.040	0.039
V. Confidants (D)	304	433	0.078	7.631	8	288	22	0.005	-0.043	0.031

Legend:
Path = average path, #C = number of components
LC = largest component, D. = diameter, RC. = reciprocity ratio
#S = number of students, #L = number of links, A. = assortativity
Trans. = transitivity

Leaving aside the special case that M.school is, the average path and –as a consequence– betweenness centrality for the confidant networks varies considerably across the schools (as seen in the table above). Reciprocity in this case may not always hold and so this has to be taken into account when interpreting the similar values regarding closeness centralities (as seen in the table below).

Table 3.14: Betweenness and eigenvector network centralities (confidants). Source: author.

	Betweenness	Eigenvector
J. Confidants (D)	3.696 (10.674), 0	0 (0.177), 0
K. Confidants (D)	93.092 (252.587), 0	0.067 (0.151), 0
M. Confidants (D)	0.364 (1.055), 0	0 (0), 0
R. Confidants (D)	176.888 (444.134), 0	0.027 (0.1), 0
V. Confidants (D)	153.424 (379.581), 0	0.054 (0.148), 0

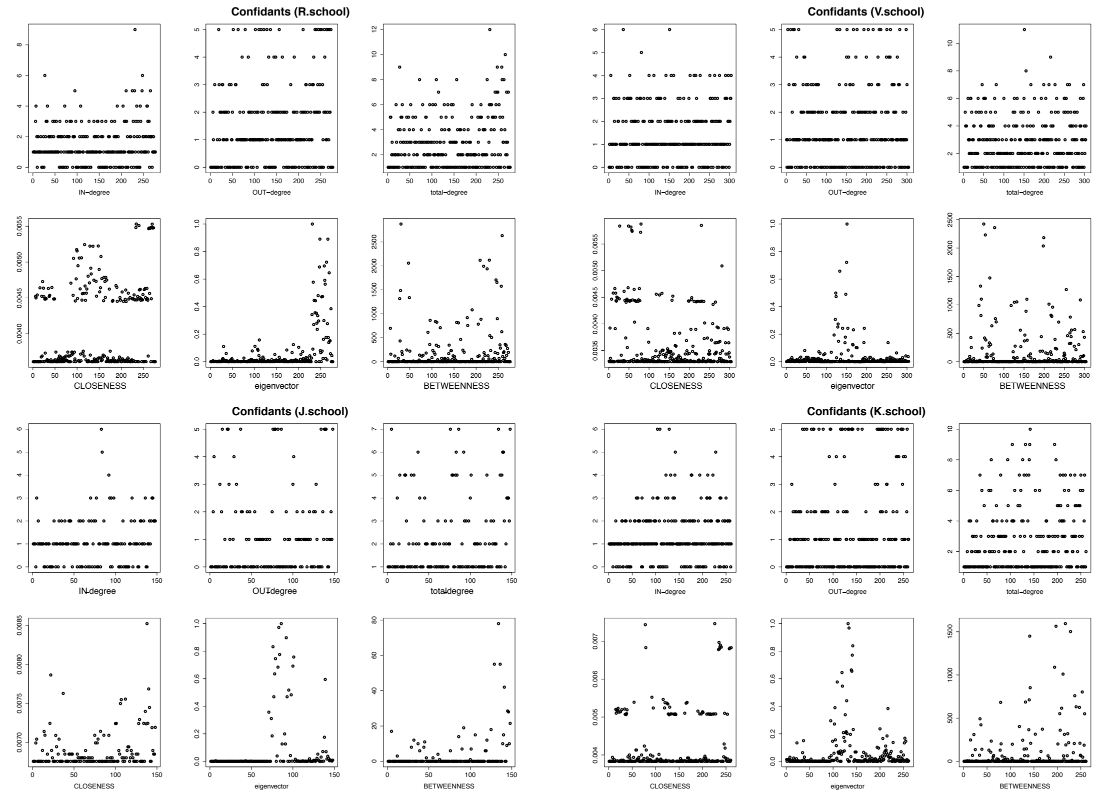
Format: Average centrality measure (standard deviation), median

Table 3.15: Closeness network centralities (confidant friends). Source: author.

	Closeness (ALL)	Closeness (IN)	Closeness (OUT)
J. Confidants (D)	0.017 (0.007), 0.021	0.007 (0), 0.007	0.007 (0), 0.007
K. Confidants (D)	0.028 (0.008), 0.03	0.004 (0), 0.004	0.004 (0.001), 0.004
M. Confidants (D)	0.036 (0.004), 0.034	0.031 (0.001), 0.031	0.032 (0.002), 0.03
R. Confidants (D)	0.028 (0.008), 0.03	0.004 (0.001), 0.004	0.004 (0.001), 0.004
V. Confidants (D)	0.042 (0.009), 0.044	0.004 (0), 0.003	0.004 (0.001), 0.003

Format: Average centrality measure (standard deviation), median

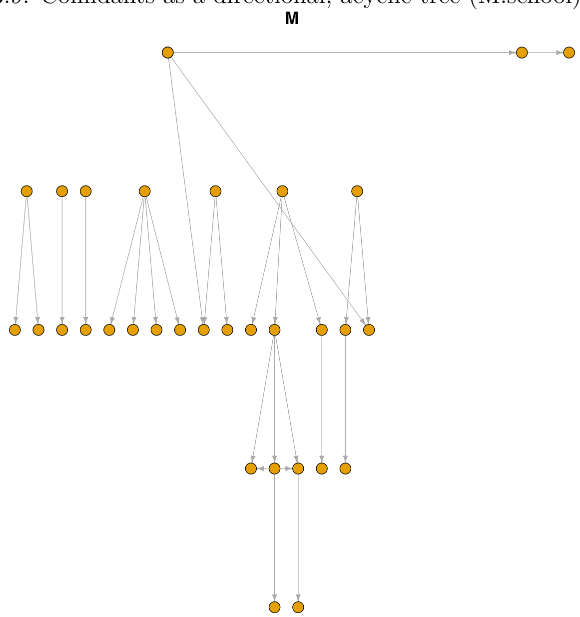
Figure 3.8: Confidants (network coefficients across activities) Source: author.



The figure above depicts the configuration of the confidant networks according to its coefficients. In and out degrees are expectedly rather uniform, with most scoring low and few scoring high. Closeness in schools R., K. and V. have clusters concentrated at middle-range values, suggesting the presence of chains among students. Eigenvector and betweenness are similar across most schools, with the exception of R. school where the eigenvector values are rather right-skewed.

In the case of confidant networks there cannot be self-cycles (i.e it is impossible to confide to yourself). Nevertheless there may be cycles (also known as social circuits in the SNA literature), of the following type: A→B, B→...→A). The only school that does not contain one is M.school. For this reason it is the only network plotted using the [Reingold and Tilford, 1981] algorithm, which tend to work best with acyclic networks (i.e. without loops). Another option would have been to apply the [Sugiyama et al., 1981] approach, also referred to in the literature as layered approach, which is also well suited for acyclic graphs.

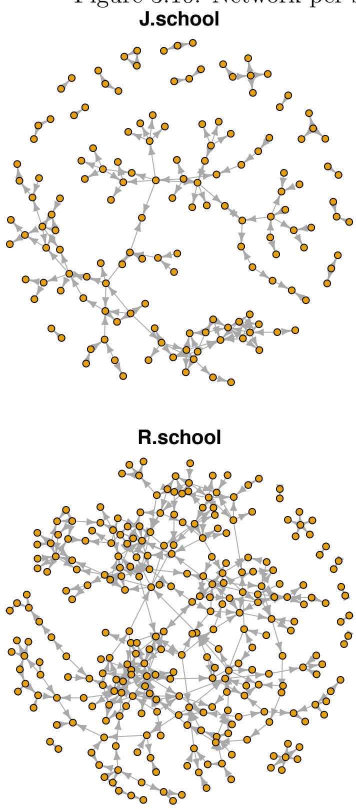
Figure 3.9: Confidants as a directional, acyclic tree (M.school) Source: author.



The confidant relationships of M.school can thus be thought as a directed tree-like structure, which is also known as a singly-connected graph [Barber, 2012]. In this case its isomorphs (i.e. mathematically identical) structures are drawn as such to facilitate understanding the overall order and symmetry –with no crossings to comply with the specification of a fully or near planar graph (i.e. embeddable).

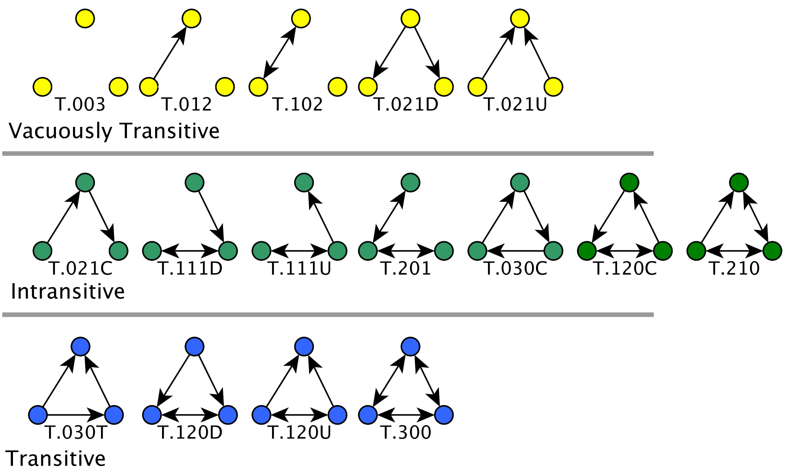
As the other networks contain cycles, these are drawn using the same approach (Fruchterman-Reingold) as used for the other friends network (figure 3.6, page 37). Removing cycles would have changed the network structure and drawing directed graphs containing loops as trees is not a computationally straightforward task. This is due to the exponential multiple ordering possibilities, imposed by graphs with multiple directed cycles, when attempting to plot tree-like structures that are usually best understood and visualised hierarchically [Oliver Bastert, 2003]. That is, plotting path and/or cycle census is a problem with a complexity given as non-deterministic polynomial time (NP) complete problems, which require searching for heuristics-based solutions [Butts, 2007, Cormen et al., 2001].

Figure 3.10: Network per school, across activities (confidants) Source: author.



Triads (i.e. every triple of vertices A, B, C) are important structures in social network analysis due to the great variety of actual and potential interactions permissible by them. There is also experimental evidence supporting the importance of triads both in self-reporting networks and patterns of social interaction [Brashears and Quintane, 2015]. Every possible type of triad in a di-graph has been classified by [Davis and Leinhardt, 1967] and the figure below depicts them.

Figure 3.11: All 16 types of di-graph triads. Source: [Doran et al., 2013]



Letters indicate the orientation of the asymmetry: Down, Up, Transitive, Cyclic. Each number in the triadic census correspond to the total count of the following:

- mutual dyads (i.e. person A nominates person B and that is reciprocated);
- asymmetric dyad (i.e. person A nominates person B but not the opposite);
- null dyads (i.e. person A does not nominate person B and vice-versa).

Thus within triads there are dyad sub-structures and these are typically well explained by assumptions of dependence and inter-dependence [Wasserman and Faust, 1994]. Theoretically social networks are not expected to sustain a large variety of triads due to the presence of transitive (closure-inducing) behaviours.

Triads of the 003 type are also known as empty graphs (i.e. absence of any link between nodes), whilst triads of the 300 type are also known as complete graphs (i.e. presence of every possible link between nodes). Such latter structure may occur within only subsets of the total network (i.e. sub-graphs) and these are referred to as cliques in the literature [David and Jon, 2010]. Hence the triads shown in the previous figure can be considered intermediary structures that mediate the progressive transition from an empty to a complete graph.

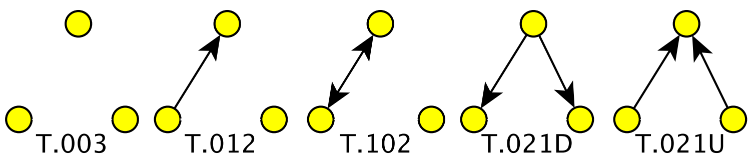
Table 3.16: Summary of vacuously transitive di-graph (census) triads. Source: author.

	003	012	102	021D	021U
J. Confidant Triads (D)	484992	22165	555	160	70
K. Confidant Triads (D)	2612703	84207	986	410	195
M. Confidant Triads (D)	4620	806	0	18	2
R. Confidant Triads (D)	3213193	98763	3945	295	247
V. Confidant Triads (D)	4510627	120696	3849	327	251

column notation: MAN (mutual, asymmetric, null)

The table above and the figure below depict the triads that are vacuously transitive (i.e. do not contain a directed path, such as A→B→C). These include dyadic only configurations, such as 012 (i.e. a single directed link), 102 (i.e. mutual link), 021D (i.e. the so-called out-star) and also the 021U setup (i.e. the so-called in-star, indicating node popularity) [21]. Understanding these sub-configurations, specially the latter two, can facilitate the explanation of the tendencies towards centralisation and transitivity in a directed network [22].

Figure 3.12: Vacuously transitive di-graph triads. Source: [Doran et al., 2013]



[21] Expansion or contraction potential [Robins et al., 2001]. These are due to resembling star shapes when incoming receivers or outgoing senders links are drawn [Cranmer et al., 2014].
[22] The approach using such specifications is discussed on page 49 and illustrated on page 64.

In and out stars can be combined into alternating stars, which represent the correlation between in and out degrees[23] –yet relying on these is typically difficult to account in terms of proposing a generative network model[24] [Wang et al., 2013]. In other words, in a given network it is more common to find individuals behaving more closely to either an in-star or out-star, rather than alternating between.

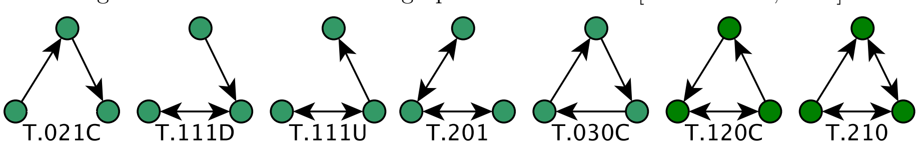
Configurations of non-transitive confidant triads are shown in the figure and table below. Intransitivity typically denotes the presence of an intermediary, or middleman, that serves to support indirect interactions. These configurations are helpful to identify potential gatekeepers (i.e. individuals that can control how information flow to some parts of a network), therefore creating some degree of social power and/or influence with regards to such individuals to allow or withhold information. As this is a confidant network, it is somewhat surprising to observe such a large number of such configurations. Yet the triad census alone is insufficient to differentiate whether these include betrayals (i.e. sharing with others of what should not) along with mutually exclusive sharing sequences.[25] [26]

Table 3.17: Summary of non-transitive di-graph (census) triads. Source: author.

	021C	111D	111U	201	030C	120C	210
J. Confidant Triads (D)	104	5	13	0	0	0	0
K. Confidant Triads (D)	450	4	14	0	3	1	0
M. Confidant Triads (D)	8	0	0	0	0	0	0
R. Confidant Triads (D)	465	40	46	2	4	1	2
V. Confidant Triads (D)	452	31	38	2	2	0	0

column notation: MAN (mutual, asymmetric, null)

Figure 3.13: Non-transitive di-graph triads. Source: [Doran et al., 2013]



[23] i.e. Number of incoming links (in-degree) and number of outgoing links (out-degree).
[24] i.e. a model that does help to re-generate and hence explain empirical network observations.
[25] i.e. paths, of different lengths, of the type A→B→C that are unique across the network.
[26] Such as the 030C cyclic triad: no two-way ties, 3 one-way ties, 0 absent ties [?].

As betweenness centrality is an important measure of how prevalent non-transitive activities are within each school, the concept of structural holes (i.e. mechanisms for brokering social capital) amongst students [Burt, 1995, 2004] may be useful both in the context of directed and non-directed networks. This is because individuals functioning as bridges between different groups may reveal critical bottlenecks in important student activities for learning. In this context, centrality measures can be helpful to understand –and perhaps predict– drop-outs and low engagement, across and within the activities students are engaged with [27].

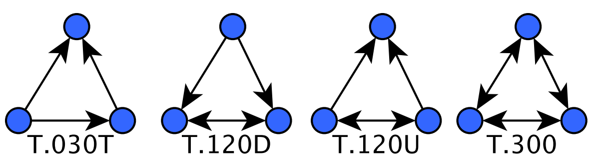
The table and figure below depict the transitive triads in the confidant networks. These tend to be associated with higher levels of homophily (i.e. greater likelihood to be linked to others with similar properties as oneself) and hence social relationships that are perceived as being strong ties. Given the generally low structural assortativity reported for confidants within schools (Table 3.13 on page 39), the census for this type of triads is expectedly rather low in terms of reported totals[28].

Table 3.18: Summary of transitive di-graph (census) triads. Source: author.

	030T	120D	120U	300
J. Confidant Triads (D)	13	2	1	0
K. Confidant Triads (D)	28	2	1	0
M. Confidant Triads (D)	2	0	0	0
R. Confidant Triads (D)	27	5	5	0
V. Confidant Triads (D)	25	1	3	0

column notation: MAN (mutual, asymmetric, null)

Figure 3.14: Transitive types of di-graph triads. Source: [Doran et al., 2013]



[27] The discussion on the relevance of these concepts is resumed in greater detail on page 39.
[28] Peripheral nodes in subgroups are often in 300-type triads, displaying high transitivity.

4. Final Considerations

The aim of this research report is to preliminary describe the network properties of five surveyed schools in India, based on the structural aggregation of the perception from each individual surveyed student. This included networks differentiated by confidants, best and other friends, regarding the following activities: studying, playing / chatting, teamwork, commuting and visiting after school.

In order to allow the analysis of networks, this document also has a detailed description of the necessary corrections and processing, up to the current date (15th December 2015), based on the original dataset. The data has been described in terms of composition (section, class) and properties regarding the following:

- directionality (symmetrical for undirected graphs, asymmetrical for di-graphs);
- centrality measures (closeness, betweenness, eigenvector for all networks);
- triad census (transitive, non-transitive, vacuously transitive configurations)[1];
- path, components, diameter, reciprocity, assortativity and transitivity.

The next section has a discussion of potential future work envisioned while preparing this report and discussing with the VAEIS project members. The document concludes with supplementary material in the appendix and bibliography.

[1] These are for the directed network (i.e. confidants). The triad census for the non-directed networks (i.e. best and other friends), with four possibilities, are in section 5.4 (pages 54, 55).

4.1 Future Work

There are various potential future work developments stemming from this report, which is focused on describing the collective structural properties of the student networks according to individual students' perceptions. The results so far include the aggregation of endogenous and self-organising mechanisms that are partially influenced by a combination of properties, which have been described: centrality measures, reciprocity, homophily, components, assortativity and transitivity [2].

The first follow-up from this report could be the integration of the network data with the other other sections from the survey administered to the Indian students –which is a task still pending the data corrections by the MVF. These include data about the student's household, teachers, headmaster, exercise sheets, geographical distances and academic achievements. This would allow a finer grain description of the network properties and sub-structures, as then multiple and more detailed criteria for sub-setting could be chosen to to understand networks within and across schools.

The VAEIS dataset is to date (15th December 2015) still being corrected and transcribed by MFV. Another aspect that could be taken into account, once the data collection and correction process are completed, is to describe and control for the geographical distance between students. This can be done by using the geographical location of households, schools and villages as additional information for each of the collected networks [3]. At the time of writing this report, this information was not yet available, so this is another suggestion for future work.

[2]These have been discussed in detail throughout section 3: "Research Findings" (page 21).
[3] e.g. [Preciado et al., 2012] is a similar approach to social structures and physical distances.

The second follow-up would be testing the network and individual properties related to academic performance. Both in terms of individual students and groups, according to criteria that is relevant to VAEIS. This would go beyond providing descriptive statistics, as tests would be done regarding the potential effects of network structure and individual attributes. Such type of hypotheses would be specifications based on Exponential Random Graphs Models (henceforth ERGM) [Robins et al., 2007]. This approach has extensions for cross-sectional (i.e. single wave, as in this report) and longitudinal (i.e. multiple data waves, which ought to be collected). As a network is rarely explained only by just one property, to understand and infer its formation process, different properties can be evaluated together in an ERGM model.

The best friends network can be analysed in the future as a fully checked undirected activity network (i.e. A→B and B→A must necessarily hold in the survey). This would be a much more stringent condition to be met if compared to the symmetrised networks described in table 2.7 (see page 20). Such approach would likely to reduce considerably some school networks, both due to missing data and genuine lack of reciprocity. For such drastic change in the network structure, other factors would be worthwhile to take into account such as student achievements as an individual control parameter and comparable measure.

Core-periphery structures could also be tested, following the highlights briefly discussed with the aid of figures 3.2 and 3.6 –focused on the network structure of best and other friends (respectively on pages 31 and 37). These include, e.g. questions as to whether the presence of 29 unique teachers in “cover.prn” in different or same sections and classes is related to academic achievement, when the structures of the students social networks are simultaneously accounted for.

Results from such type of hypotheses may provide insights on the importance of self-generated networks in relation to the school-imposed structures (for example teacher allocation and classroom layout, as seen in figure 5.10 on page 76).

In section 3.2: "Best Friends across Activities" (page 31), the sensitivity-based approach to the network visualisation has been mentioned as an alternative to transform the original structure into a simplified skeleton (i.e. sparser structure) that still captures the overall structural properties. This could facilitate the core-periphery (or core-satellite) approach to understand these networks, which consist of having one or more dense cores and a variety of fairly disconnected nodes. Such network structures typically contain "*a small sub-set of the total population participating more actively than the rest*" [Cummings and Cross, 2003].

Therefore it would be reasonable to enunciate hypotheses regarding potential better academic performances with regards to these two network structures. Given the nature of this case study , the immediate networks of relevance would be: study and class teamwork. The remaining activities (i.e. play and chat, commute, visiting after school and sharing personal problems) could also be considered in hypotheses either regarding overlapping or disjoint structures altogether[4]. As a consequence of carrying this suggestion out, there would also be an assessment of students belonging to peripheral network structures, thus shedding light on:

- whether there are significant academic achievement differences between individuals that belong to one structure or another (i.e. core-periphery)[5].
- whether the presence of different school triads amongst confidants, reported in section 3.11 on page 43, is a likely result due to popularity or transitivity.

[4]I.e. sub-sets that are present in two or more networks but not in the remaining activities. For a complete list, please refer to sections 5.27 and 5.28, respectively on pages 67 and 70.

[5] A more detailed discussion about such models [Borgatti and Everett, 2000].

5. Appendix

5.1 Network Survey Section (Other Friends)

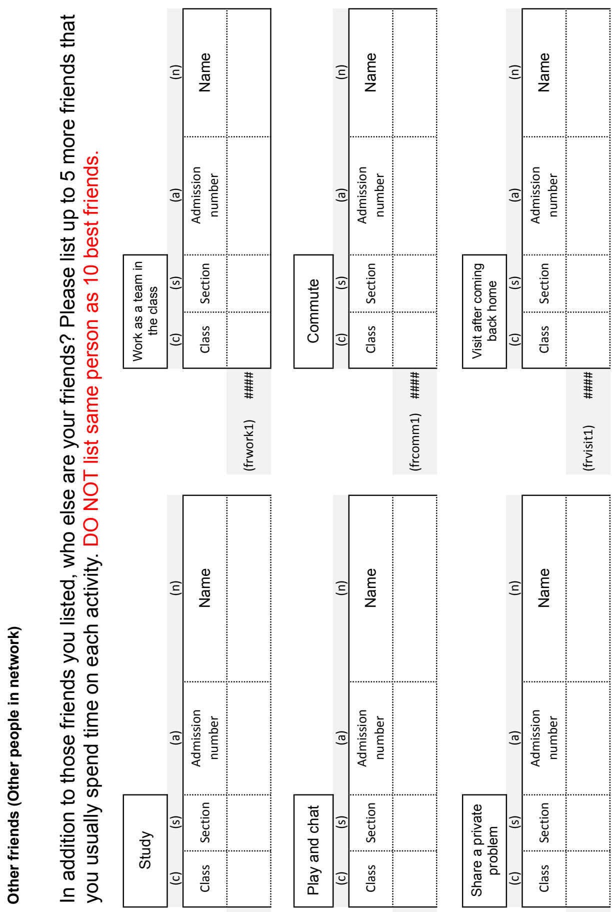


Figure 5.15: Network survey questionnaire: Others. Source: VAIES project.

5.2 Network Survey Section (Best Friends)

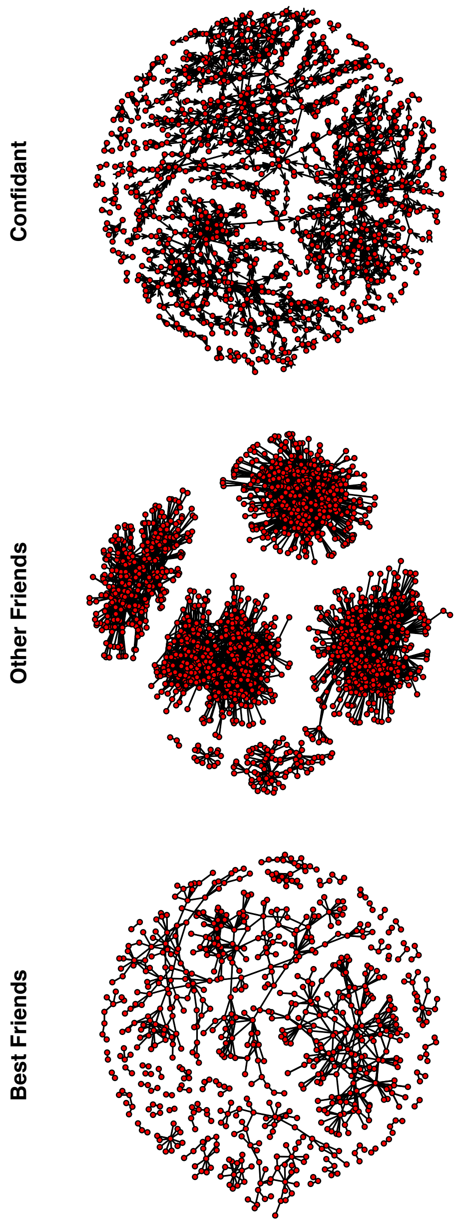
Best friends (Network, form 1)

1. You are given a list of all student names in relevant grades in this school. Please write your best friends' names and student numbers. Both boys and girls can be listed. Best friends are the ones you trust even in times of trouble.				2. Do you do the following together on most of school days with this friend? Choose Yes ☑ if you do, and give how often.								
(c)	(s)	(a)	(n)	(sy)	(sf)	(py)	(pf)	(ty)	(cy)	(vy)	(vf)	(hy)
Class	Section	Admission number	Name	Study		Play and chat		Work as a team in class	Commute	Visit each other after coming back home		Share a personal problem
				yes/no	frequency in 1 week	yes/no	frequency in 1 week	yes/no	yes/no	yes/no	frequency in 1 week	yes/no
				□Yes	□less than once □1-3 times □4+ times	□Yes	□less than once □1-3 times □4+ times	□Yes	□Yes	□Yes	□less than once □1-3 times □4+ times	□Yes
				□Yes	□less than once □1-3 times □4+ times	□Yes	□less than once □1-3 times □4+ times	□Yes	□Yes	□Yes	□less than once □1-3 times □4+ times	□Yes

Figure 5.16: Network survey questionnaire: Best Friends. Source: VAIES project.

5.3 Overall Structure: Best, Other, Confidant

Figure 5.17: Plain networks: best, other and confidant friends. Source: author.



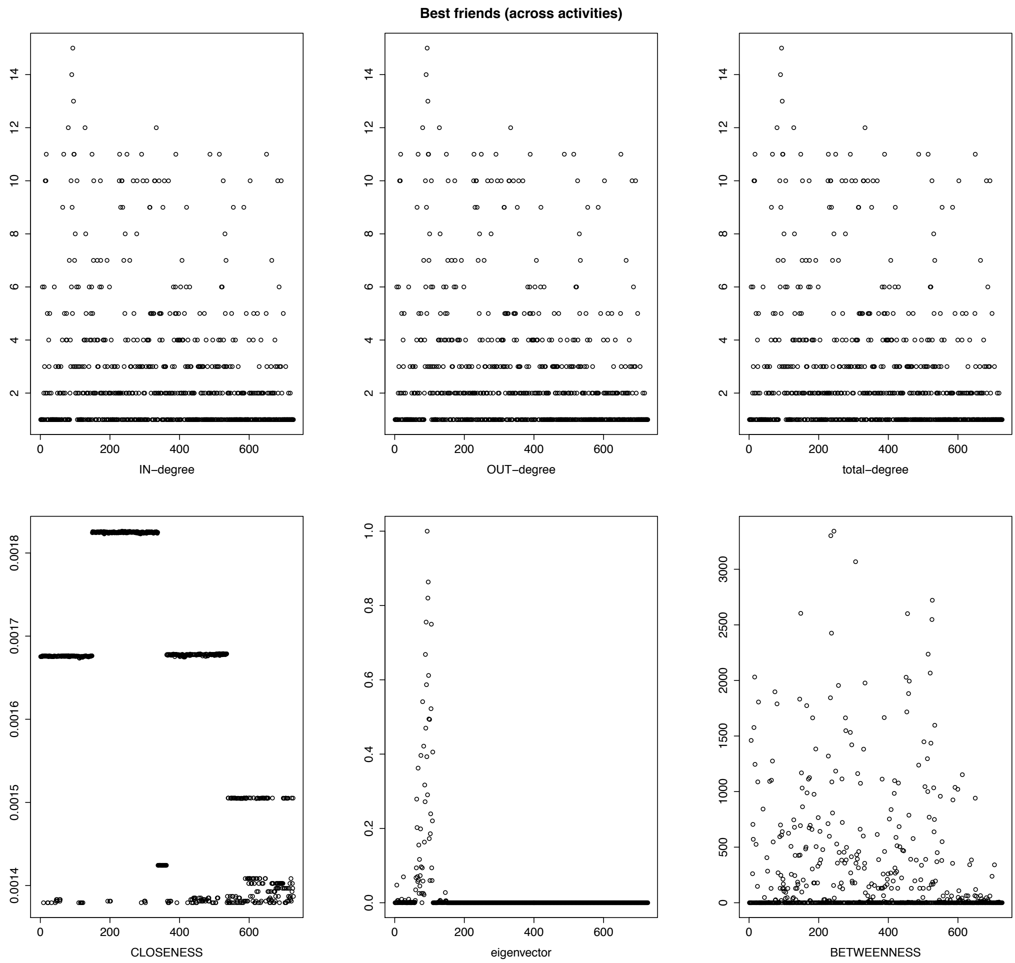


Figure 5.18: Degrees and centralities (best friends across schools and activities)
Source: author.

Table 5.19: Triad census (undirected best friends). Source: author.

	0	1	2	3
J. Best Friends Triads (U)	495003	33280	995	118
K. Best Friends Triads (U)	1036441	52193	1115	87
M. Best Friends Triads (U)	1498	661	118	23
R. Best Friends Triads (U)	824923	37356	591	54
V. Best Friends Triads (U)	1109759	32932	433	11

Four types of triads for undirected networks:
no link, one link, two links, or all three links.

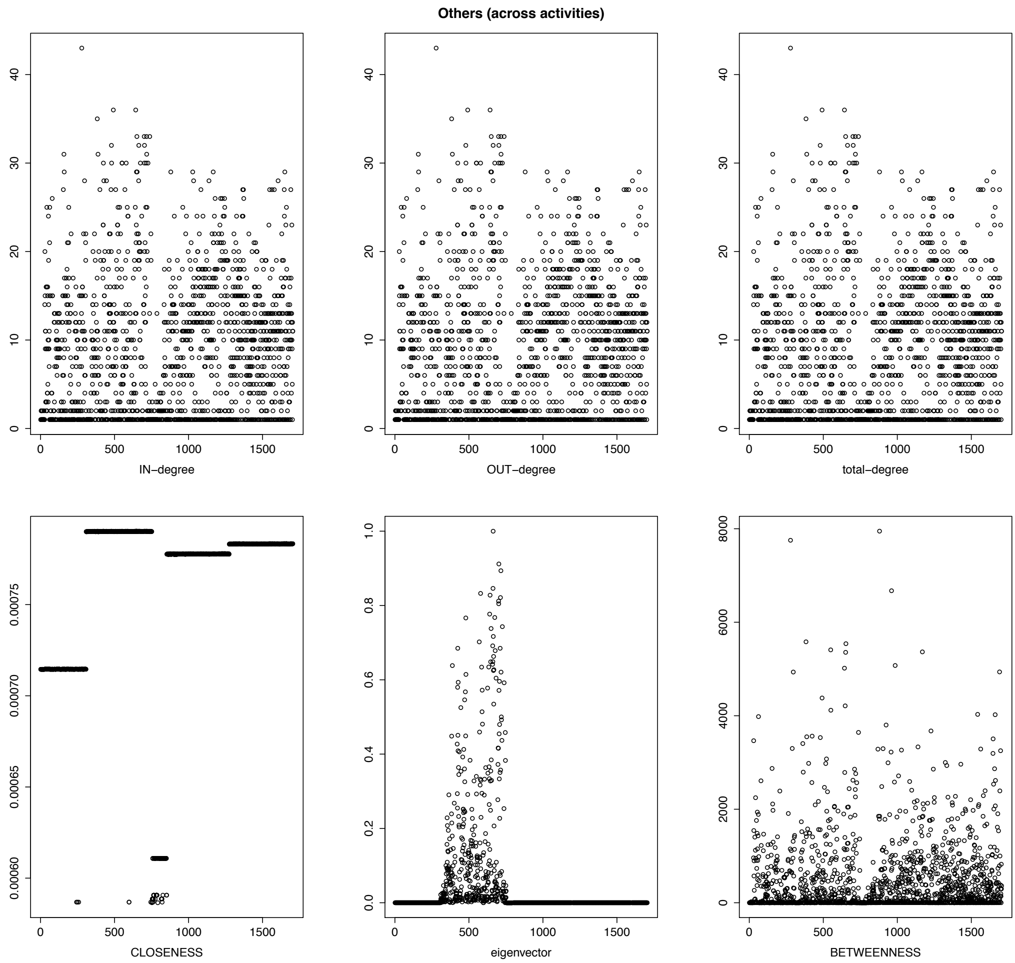


Figure 5.19: Degrees and centralities (others across schools and activities). Source: author.

Table 5.20: Triad census (undirected other friends) Source： author

	0	1	2	3
J. Other Friends Triads (U)	4650343	303103	10570	1099
K. Other Friends Triads (U)	13765141	794482	25922	2445
M. Other Friends Triads (U)	135370	11342	701	27
R. Other Friends Triads (U)	11657620	755929	20422	2369
V. Other Friends Triads (U)	12386703	841428	21003	1961

Four types of triads for undirected networks:
no link, one link, two links, or all three links.

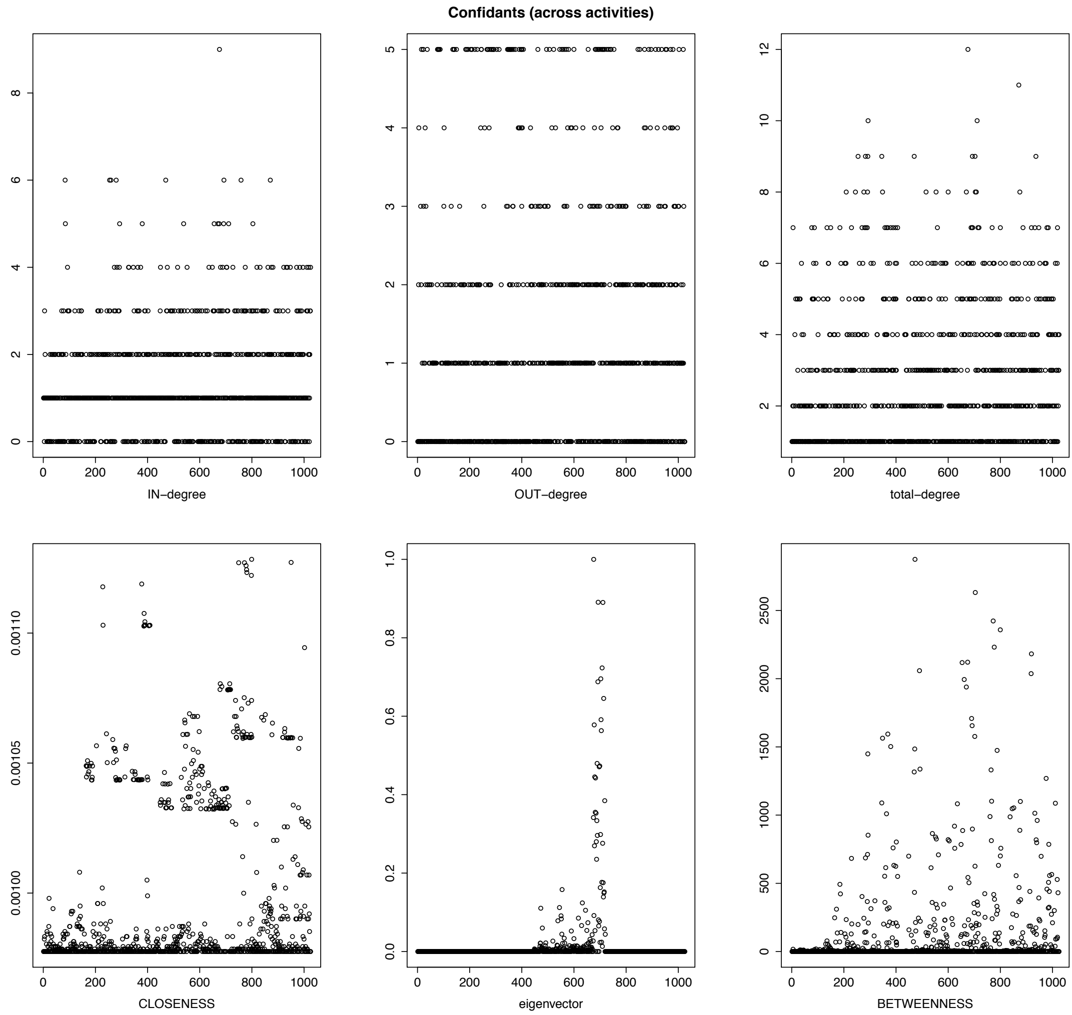


Figure 5.20: Degrees and centralities (confidants across schools and activities). Source: author.

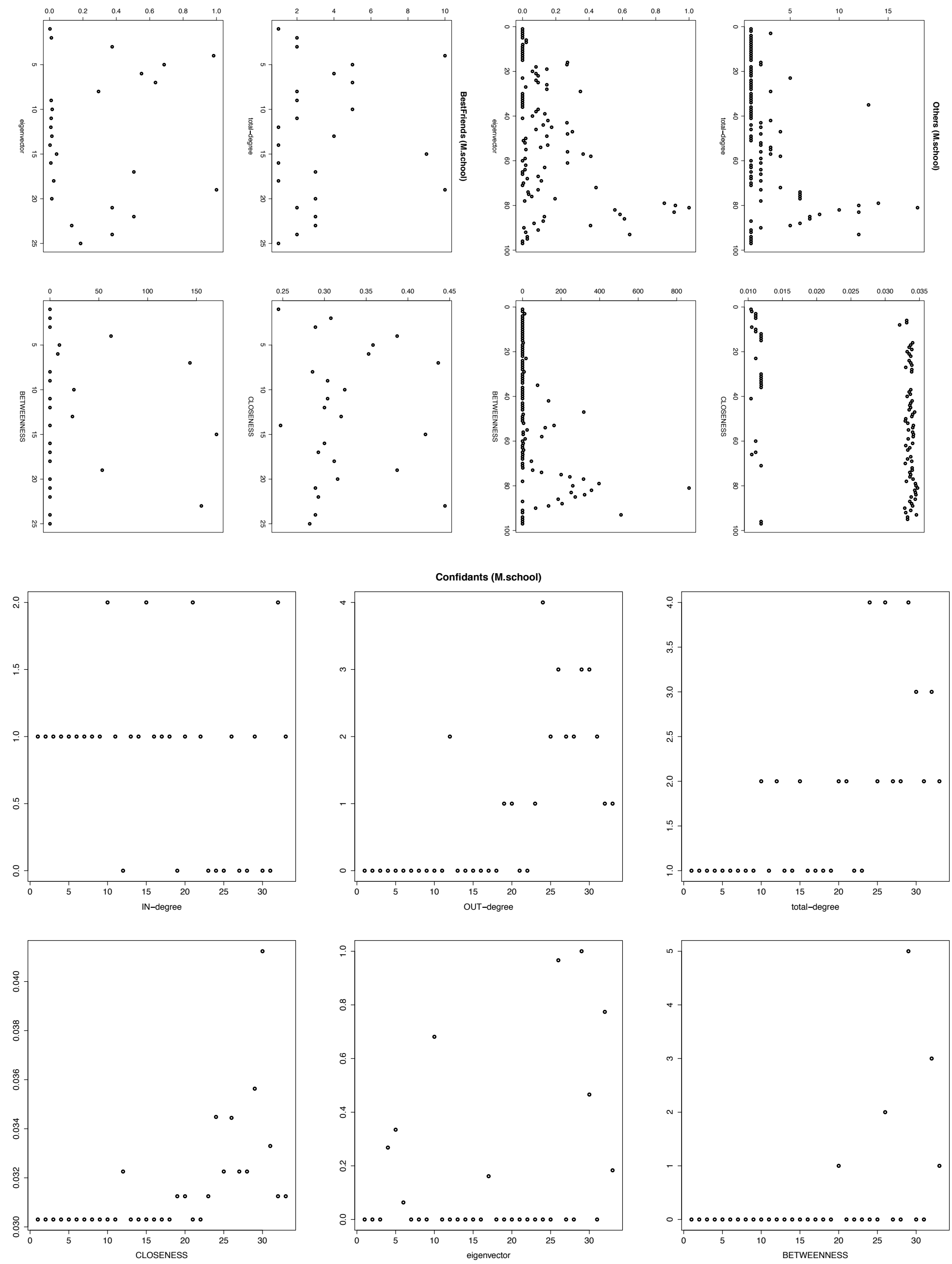
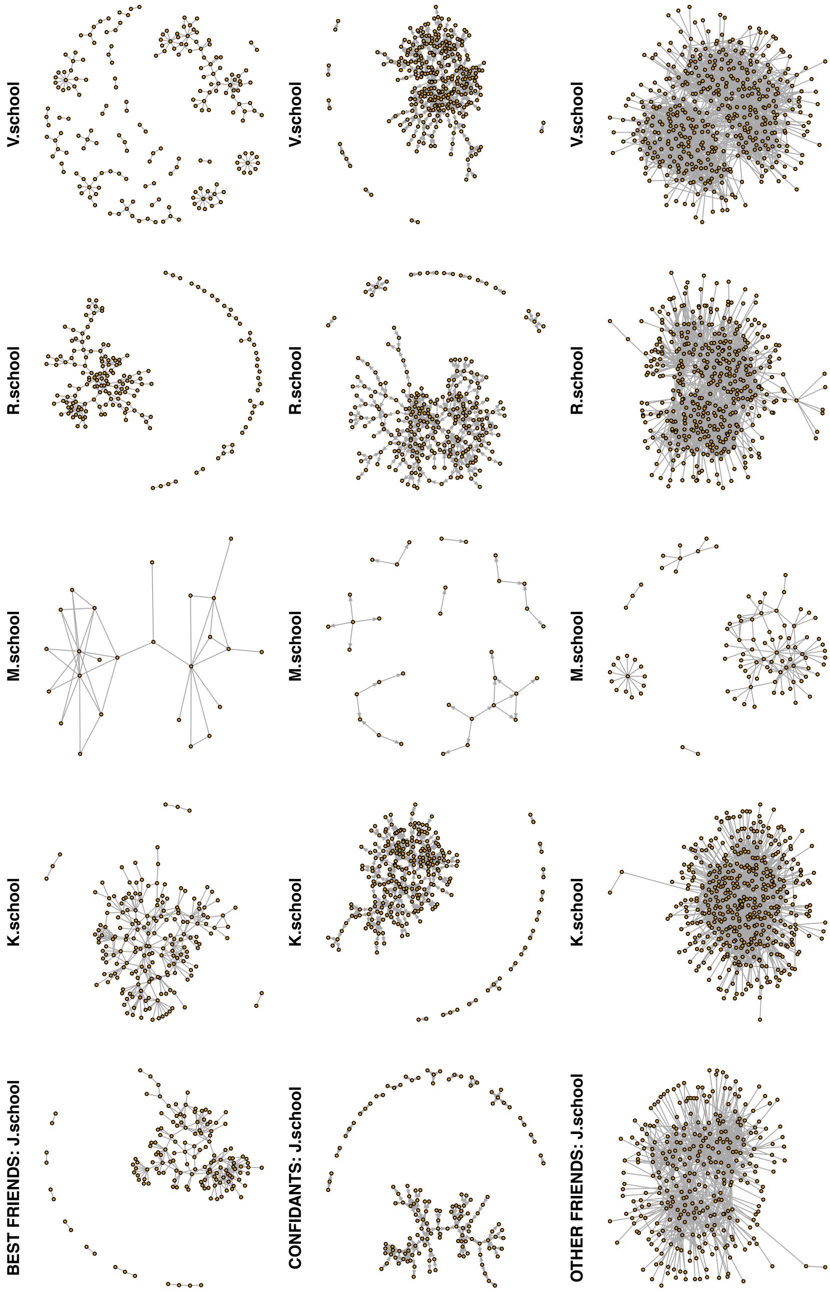


Figure 5.21: Degrees and centralities (confidants, best and others: M.school). Source: author.

5.4 Overall Network Structure, per School

Figure 5.22: Plain networks: best, other and confidant friends. Source: author.



5.5 Summary of Best, Others and Confidants

[1]

Table 5.21: Network size per school, section and class (other friends). Source: author.

	school	section	class	8	9	10	Total
1							
2	J.school	a (telugu)		140	139	351	630
3		b (english)		264	289	142	695
4		c		0	0	0	0
5	K.school	a (telugu)		467	1062	313	1842
6		b (english)		233	762	469	1464
7		c		0	0	0	0
8	M.school	a (telugu)		7	125	19	151
9		b (english)		0	0	0	0
10		c		0	0	0	0
11	R.school	a (telugu)		335	425	289	1049
12		b (english)		296	615	343	1254
13		c		261	0	0	261
14	V.school	a (telugu)		249	334	313	896
15		b (english)		342	335	482	1159
16		c		286	309	0	595

[1] For the full subset composition per activity, refer to the table in section 5.28 (page 70).

Table 5.22: Network size per section per school (confidants). Source: author.

			class	8	9	10	Total
1	school	section					
2	J.school	a (telugu)		18	18	50	86
3		b (english)		22	42	17	81
4		c		0	0	0	0
5	K.school	a (telugu)		85	152	46	283
6		b (english)		33	120	62	215
7		c		0	0	0	0
8	M.school	a (telugu)		2	24	2	28
9		b (english)		0	0	0	0
10		c		0	0	0	0
11	R.school	a (telugu)		52	53	54	159
12		b (english)		50	103	53	206
13		c		39	0	0	39
14	V.school	a (telugu)		27	67	54	148
15		b (english)		63	42	74	179
16		c		52	55	0	107

Table 5.23: Summary of classes and sections (all networks). Source: author.

	#classes	#sections
J. Best Friends (U)	3	2
K. Best Friends (U)	3	2
M. Best Friends (U)	1	1
R. Best Friends (U)	3	3
V. Best Friends (U)	3	3
J. Others (U)	3	2
K. Others (U)	3	2
M. Others (U)	3	1
R. Others (U)	3	3
V. Others (U)	3	3
J. Confidants (D)	3	2
K. Confidants (D)	3	2
M. Confidants (D)	3	1
R. Confidants (D)	3	3
V. Confidants (D)	3	3

Table 5.24: Summary of network statistics (all schools and networks). Source: author.

	#S	#L	Trans.	Path	#C.	LC.	D.	Den.	A.	RC.
J. Best Friends (U)	148	244	0.262	5.143	8	131	13	0.022	-0.050	1
K. Best Friends (U)	188	294	0.190	5.002	4	180	10	0.017	-0.370	1
M. Best Friends (U)	25	42	0.369	3.167	1	25	6	0.140	-0.488	1
R. Best Friends (U)	174	225	0.215	7.017	15	132	19	0.015	0.028	1
V. Best Friends (U)	191	179	0.071	5.415	28	63	15	0.010	-0.332	1
J. Others (U)	311	1060	0.238	3.507	2	308	7	0.022	-0.103	1
K. Others (U)	445	1927	0.221	3.372	2	442	7	0.020	-0.025	1
M. Others (U)	97	135	0.104	3.288	5	70	8	0.029	-0.328	1
R. Others (U)	422	1914	0.258	3.629	1	422	9	0.022	0.078	1
V. Others (U)	431	2073	0.219	3.454	1	431	7	0.022	0.052	1
J. Confidants (D)	148	167	0.123	2.284	15	107	8	0.008	0.312	0.025
K. Confidants (D)	261	351	0.089	5.639	13	234	14	0.005	0.094	0.012
M. Confidants (D)	33	28	0.176	1.316	7	10	3	0.027	0.048	0
R. Confidants (D)	277	404	0.109	7.487	10	250	25	0.005	0.040	0.039
V. Confidants (D)	304	433	0.078	7.631	8	288	22	0.005	-0.043	0.031

Legend:
#S = number of students
#L = number of links
Trans. = Transitivity
Path = average path
#C = number of components
LC = largest component
D. = diameter
Den. = Density
A. = Assortativity
RC. = Reciprocity Ratio

Table 5.25: Summary of network centralities (all networks). Source: author.

	Betweenness	Closeness	Eigenvector
J. Best Friends (U)	0.042 (0.013), 0.046	0.099 (0.203), 0.001	238.736 (484.939), 0
K. Best Friends (U)	0.075 (0.015), 0.078	0.048 (0.16), 0.006	343.053 (616.465), 0
M. Best Friends (U)	0.324 (0.054), 0.304	0.271 (0.315), 0.133	26 (51.959), 0
R. Best Friends (U)	0.017 (0.007), 0.021	0.064 (0.189), 0	300.69 (574.482), 1
V. Best Friends (U)	0.006 (0.001), 0.006	0.032 (0.118), 0	56.356 (190.943), 0
J. Others (U)	0.154 (0.019), 0.154	0.111 (0.182), 0.044	381.148 (782.894), 13.375
K. Others (U)	0.157 (0.018), 0.16	0.16 (0.212), 0.067	519.622 (951.318), 50.076
M. Others (U)	0.027 (0.01), 0.033	0.141 (0.22), 0.06	59.876 (134.919), 0
R. Others (U)	0.283 (0.042), 0.287	0.145 (0.22), 0.047	553.436 (912.947), 197.987
V. Others (U)	0.294 (0.035), 0.298	0.121 (0.182), 0.062	527.603 (740.479), 266.685

Format: Average centrality measure (standard deviation), median

Table 5.26: Summary of network centralities (confidants). Source: author.

	Betweenness	Closeness (ALL)	Closeness (IN)	Closeness (OUT)	Eigenvector
J. Confidants (D)	3.696 (10.674), 0	0.017 (0.007), 0.021	0.007 (0), 0.007	0.007 (0), 0.007	0 (0.177), 0
K. Confidants (D)	93.092 (252.587), 0	0.028 (0.008), 0.03	0.004 (0), 0.004	0.004 (0.001), 0.004	0.067 (0.151), 0
M. Confidants (D)	0.364 (1.055), 0	0.036 (0.004), 0.034	0.031 (0.001), 0.031	0.032 (0.002), 0.03	0 (0), 0
R. Confidants (D)	176.888 (444.134), 0	0.028 (0.008), 0.03	0.004 (0.001), 0.004	0.004 (0.001), 0.004	0.027 (0.1), 0
V. Confidants (D)	153.424 (379.581), 0	0.042 (0.009), 0.044	0.004 (0), 0.003	0.004 (0.001), 0.003	0.054 (0.148), 0

5.6 Illustration of Some ERGM Specification Terms

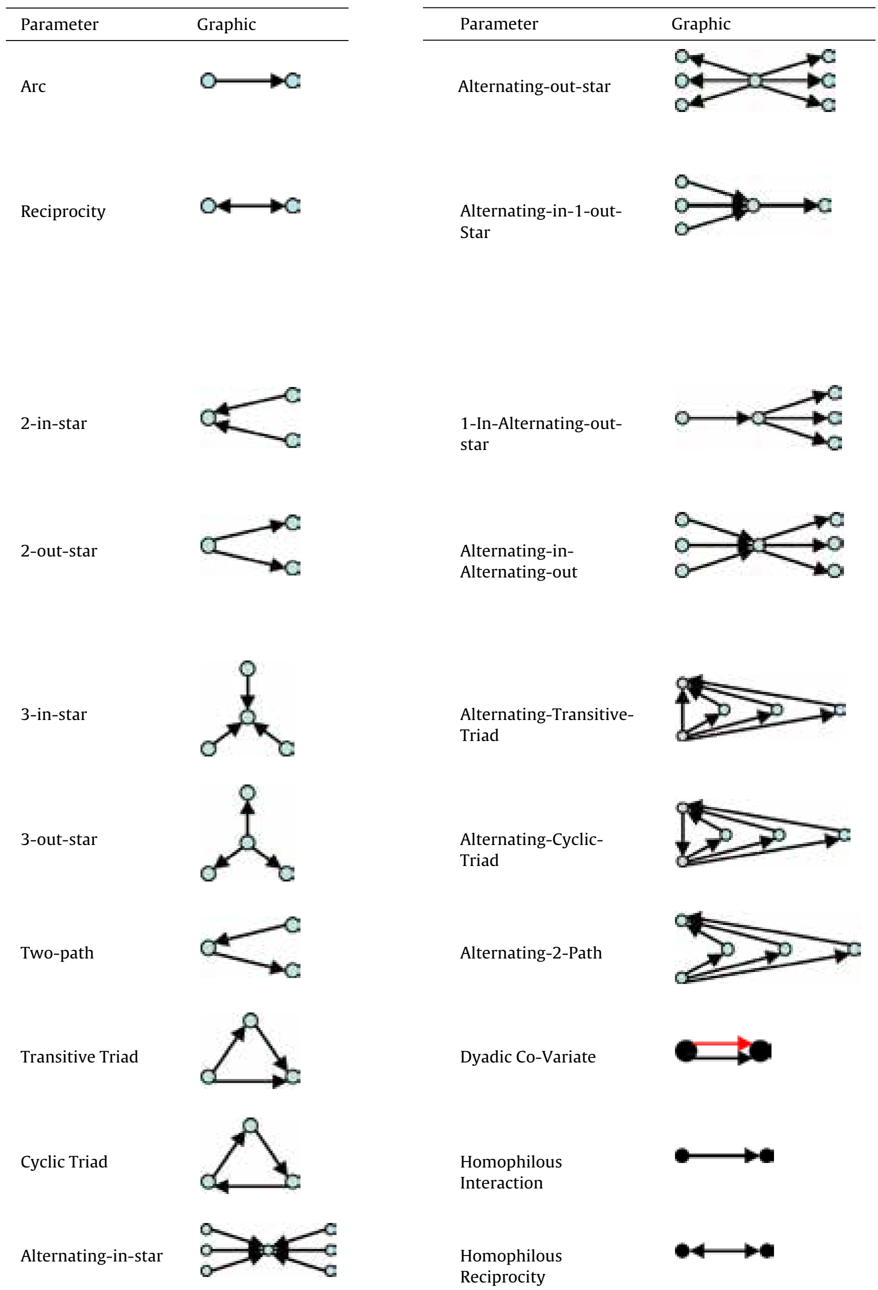


Figure 5.23: ERGM terms illustration. Source: [Gondal and McLean, 2013]

5.7 Activity per school, Section, Class (Best)

				class	8	9	10	Total
1	network	school	section					
2	class_team	J.school	a (telugu)		39	10	57	106
3			b (english)		89	53	15	157
4			c		0	0	0	0
5		K.school	a (telugu)		77	94	54	225
6			b (english)		37	124	34	195
7			c		0	0	0	0
8		M.school	a (telugu)		0	47	0	47
9			b (english)		0	0	0	0
10			c		0	0	0	0
11		R.school	a (telugu)		21	25	29	75
12			b (english)		22	66	58	146
13			c		14	0	0	14
14		V.school	a (telugu)		39	32	28	99
15			b (english)		31	8	22	61
16			c		22	3	0	25
17	commute	J.school	a (telugu)		39	10	57	106
18			b (english)		89	53	15	157
19			c		0	0	0	0
20		K.school	a (telugu)		77	94	54	225
21			b (english)		37	124	34	195
22			c		0	0	0	0
23		M.school	a (telugu)		0	47	0	47
24			b (english)		0	0	0	0
25			c		0	0	0	0

26		R.school	a (telugu)	21	25	29	75
27			b (english)	22	66	58	146
28			c	14	0	0	14
29		V.school	a (telugu)	39	32	28	99
30			b (english)	31	8	22	61
31			c	22	3	0	25
32	play_chat	J.school	a (telugu)	39	10	57	106
33			b (english)	89	53	15	157
34			c	0	0	0	0
35		K.school	a (telugu)	77	94	54	225
36			b (english)	37	124	34	195
37			c	0	0	0	0
38		M.school	a (telugu)	0	47	0	47
39			b (english)	0	0	0	0
40			c	0	0	0	0
41		R.school	a (telugu)	21	25	29	75
42			b (english)	22	66	58	146
43			c	14	0	0	14
44		V.school	a (telugu)	39	32	28	99
45			b (english)	31	8	22	61
46			c	22	3	0	25
47	study	J.school	a (telugu)	39	10	57	106
48			b (english)	89	53	15	157
49			c	0	0	0	0
50		K.school	a (telugu)	77	94	54	225
51			b (english)	37	124	34	195
52			c	0	0	0	0
53		M.school	a (telugu)	0	47	0	47
54			b (english)	0	0	0	0
55			c	0	0	0	0

56		R.school	a (telugu)	21	25	29	75
57			b (english)	22	66	58	146
58			c	14	0	0	14
59		V.school	a (telugu)	39	32	28	99
60			b (english)	31	8	22	61
61			c	22	3	0	25
62	visit	J.school	a (telugu)	39	10	57	106
63			b (english)	89	53	15	157
64			c	0	0	0	0
65		K.school	a (telugu)	77	94	54	225
66			b (english)	37	124	34	195
67			c	0	0	0	0
68		M.school	a (telugu)	0	47	0	47
69			b (english)	0	0	0	0
70			c	0	0	0	0
71		R.school	a (telugu)	21	25	29	75
72			b (english)	22	66	58	146
73			c	14	0	0	14
74		V.school	a (telugu)	39	32	28	99
75			b (english)	31	8	22	61
76			c	22	3	0	25

Table 5.27: Activity per school, section, class (best friends). Source: author.

5.8 Activity per School, Section, Class (Others)

				class	8	9	10	Total
1	network	school	section					
2	commute	J.school	a (telugu)		13	22	53	88
3			b (english)		27	49	21	97
4			c		0	0	0	0
5		K.school	a (telugu)		68	168	53	289
6			b (english)		40	126	83	249
7			c		0	0	0	0
8		M.school	a (telugu)		3	14	2	19
9			b (english)		0	0	0	0
10			c		0	0	0	0
11		R.school	a (telugu)		59	82	46	187
12			b (english)		27	108	53	188
13			c		46	0	0	46
14		V.school	a (telugu)		39	53	59	151
15			b (english)		64	60	82	206
16			c		57	47	0	104
17	play and chat	J.school	a (telugu)		35	26	58	119
18			b (english)		57	57	25	139
19			c		0	0	0	0
20		K.school	a (telugu)		103	256	64	423
21			b (english)		46	166	102	314
22			c		0	0	0	0
23		M.school	a (telugu)		0	30	6	36
24			b (english)		0	0	0	0
25			c		0	0	0	0

26		R.school	a (telugu)	75	79	65	219
27			b (english)	95	128	85	308
28			c	60	0	0	60
29		V.school	a (telugu)	40	65	60	165
30			b (english)	48	61	98	207
31			c	52	59	0	111
32	study	J.school	a (telugu)	49	45	105	199
33			b (english)	86	75	60	221
34			c	0	0	0	0
35		K.school	a (telugu)	148	316	87	551
36			b (english)	69	190	129	388
37			c	0	0	0	0
38		M.school	a (telugu)	2	37	7	46
39			b (english)	0	0	0	0
40			c	0	0	0	0
41		R.school	a (telugu)	95	121	81	297
42			b (english)	95	146	105	346
43			c	69	0	0	69
44		V.school	a (telugu)	86	92	97	275
45			b (english)	120	104	133	357
46			c	78	86	0	164
47	visit after school	J.school	a (telugu)	14	22	49	85
48			b (english)	25	42	15	82
49			c	0	0	0	0
50		K.school	a (telugu)	64	122	46	232
51			b (english)	36	128	65	229
52			c	0	0	0	0
53		M.school	a (telugu)	0	21	2	23
54			b (english)	0	0	0	0
55			c	0	0	0	0

56		R.school	a (telugu)	47	59	35	141
57			b (english)	31	97	36	164
58			c	31	0	0	31
59		V.school	a (telugu)	37	47	39	123
60			b (english)	39	39	68	146
61			c	41	54	0	95
62	work as a team	J.school	a (telugu)	29	24	86	139
63			b (english)	69	66	21	156
64			c	0	0	0	0
65		K.school	a (telugu)	84	200	63	347
66			b (english)	42	152	90	284
67			c	0	0	0	0
68		M.school	a (telugu)	2	23	2	27
69			b (english)	0	0	0	0
70			c	0	0	0	0
71		R.school	a (telugu)	59	84	62	205
72			b (english)	48	136	64	248
73			c	55	0	0	55
74		V.school	a (telugu)	47	77	58	182
75			b (english)	71	71	101	243
76			c	58	63	0	121

Table 5.28: Activity per school, section, class (other friends). Source: author.

5.9 Undirected Best Friends Network Sizes

	school	study	play / chat	class team	commute	visit	size
1	J.school	yes	yes	yes	yes	yes	263
2	J.school	yes	yes	yes	yes		52
3	J.school	yes	yes	yes		yes	26
4	J.school	yes	yes	yes			50
5	J.school	yes	yes		yes	yes	31
6	J.school	yes	yes		yes		9
7	J.school	yes	yes			yes	4
8	J.school	yes	yes				32
9	J.school	yes		yes	yes	yes	9
10	J.school	yes		yes	yes		18
11	J.school	yes		yes		yes	4
12	J.school	yes		yes			18
13	J.school	yes			yes	yes	8
14	J.school	yes			yes		6
15	J.school	yes					100
16	J.school		yes	yes	yes	yes	4
17	J.school		yes	yes	yes		3
18	J.school		yes	yes		yes	3
19	J.school		yes	yes			2
20	J.school		yes		yes		7
21	J.school		yes				29
22	J.school			yes	yes	yes	4
23	J.school			yes	yes		12
24	J.school			yes			19
25	J.school				yes		14
26	J.school					yes	6
27	J.school						209

28	K.school	yes	yes	yes	yes	yes	426
29	K.school	yes	yes	yes	yes		70
30	K.school	yes	yes	yes		yes	23
31	K.school	yes	yes	yes			52
32	K.school	yes	yes		yes	yes	35
33	K.school	yes	yes		yes		38
34	K.school	yes	yes			yes	9
35	K.school	yes	yes				92
36	K.school	yes		yes	yes	yes	15
37	K.school	yes		yes	yes		12
38	K.school	yes		yes		yes	8
39	K.school	yes		yes			16
40	K.school	yes			yes		8
41	K.school	yes				yes	11
42	K.school	yes					196
43	K.school		yes	yes	yes	yes	24
44	K.school		yes	yes	yes		15
45	K.school		yes	yes		yes	9
46	K.school		yes	yes			11
47	K.school		yes		yes	yes	16
48	K.school		yes		yes		21
49	K.school		yes			yes	16
50	K.school		yes				154
51	K.school			yes	yes	yes	5
52	K.school			yes	yes		6
53	K.school			yes		yes	5
54	K.school			yes			73
55	K.school				yes	yes	8
56	K.school				yes		87
57	K.school					yes	66

58	K.school						310
59	M.school	yes	yes	yes	yes	yes	48
60	M.school	yes	yes	yes	yes		5
61	M.school	yes	yes	yes			23
62	M.school	yes	yes		yes	yes	3
63	M.school	yes	yes				8
64	M.school	yes		yes			4
65	M.school	yes			yes	yes	3
66	M.school	yes				yes	2
67	M.school	yes					8
68	M.school		yes		yes		2
69	M.school						8
70	R.school	yes	yes	yes	yes	yes	235
71	R.school	yes	yes	yes	yes		93
72	R.school	yes	yes	yes		yes	49
73	R.school	yes	yes	yes			178
74	R.school	yes	yes		yes	yes	34
75	R.school	yes	yes		yes		31
76	R.school	yes	yes			yes	26
77	R.school	yes	yes				101
78	R.school	yes		yes	yes	yes	7
79	R.school	yes		yes	yes		6
80	R.school	yes		yes		yes	10
81	R.school	yes		yes			31
82	R.school	yes			yes	yes	4
83	R.school	yes			yes		15
84	R.school	yes				yes	4
85	R.school	yes					122
86	R.school		yes	yes	yes	yes	20
87	R.school		yes	yes	yes		11

88	R.school		yes	yes		yes	6
89	R.school		yes	yes			19
90	R.school		yes		yes	yes	7
91	R.school		yes		yes		16
92	R.school		yes			yes	6
93	R.school		yes				70
94	R.school			yes	yes	yes	8
95	R.school			yes	yes		5
96	R.school			yes		yes	2
97	R.school			yes			39
98	R.school				yes	yes	7
99	R.school				yes		30
100	R.school					yes	13
101	R.school						206
102	V.school	yes	yes	yes	yes	yes	185
103	V.school	yes	yes	yes	yes		102
104	V.school	yes	yes	yes		yes	15
105	V.school	yes	yes	yes			113
106	V.school	yes	yes		yes	yes	33
107	V.school	yes	yes		yes		38
108	V.school	yes	yes			yes	12
109	V.school	yes	yes				85
110	V.school	yes		yes	yes	yes	29
111	V.school	yes		yes	yes		52
112	V.school	yes		yes		yes	9
113	V.school	yes		yes			151
114	V.school	yes			yes	yes	29
115	V.school	yes			yes		60
116	V.school	yes				yes	4
117	V.school	yes					342

118	V.school	yes	yes	yes	yes	5
119	V.school	yes	yes	yes		5
120	V.school	yes	yes		yes	5
121	V.school	yes	yes			19
122	V.school	yes		yes	yes	9
123	V.school	yes		yes		12
124	V.school	yes			yes	8
125	V.school	yes				69
126	V.school		yes	yes	yes	9
127	V.school		yes	yes		11
128	V.school		yes			39
129	V.school			yes	yes	7
130	V.school			yes		63
131	V.school				yes	20
132	V.school					196

Table 5.29: Undirected network sizes (per school). Source: author.

5.10 Example of a Surveyed School: Layout

Both photos included in this report have been taken by VAEIS team members.

Figure 5.24: Example of a surveyed school: classroom layout

6. About the Author

Dr Pablo Lucas is an Assistant Professor (Lecturer) at the University College Dublin, Ireland, with research focused on computational social science, particularly experimental design using agent-based models and network analysis.

His education include reading for a Bachelor of Science in Computer Science (BSc) focused on evolutionary networks, a Master of Philosophy (MPhil) and a Doctor of Philosophy (PhD) focused on agent-based social simulation.

Bibliography

Anthonisse, J. M. (1971). The rush in a directed graph. Technical report, Stichting Mathemastisch Centrum, Amsterdam, The Netherlands.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The Diffusion of Microfinance. *Science*, 341(6144):1236498.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press, New York, NY, USA.

Bojanowski, M. and Corten, R. (2014). Measuring segregation in social networks. *Social Networks*, 39:14 – 32.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27(1):55 – 71.

Borgatti, S. P. and Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4):375 – 395.

Brashears, M. E. and Quintane, E. (2015). The microstructures of network recall: How social networks are encoded and represented in human memory. *Social Networks*, 41:113–126.

Burt, R. S. (1995). *Structural holes : social structure of competition*. Harvard University Press, Cambridge.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2):pp. 349–399.

Butts, C. T. (2007). Social network analysis with sna. *Journal of Statistical Software*, 24(6):1–51.

Butts, C. T. (2008). network: a package for managing relational data in r. *Journal of Statistical Software*, 24(2).

Chang, W. (2012). *R Graphics Cookbook*. O'Reilly Media.

Chen, I. and Yang, C. (2010). Visualization of social networks. In *Handbook of Social Network Technologies and Applications*, pages 585–610.

Cormen, T. H., Stein, C., Rivest, R. L., and Leiserson, C. E. (2001). *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition.

Cranmer, S. J., Heinrich, T., and Desmarais, B. A. (2014). Reciprocity and the structural determinants of the international sanctions network. *Social Networks*, 36:5 – 22. Special Issue on Political Networks.

Crnovrsanin, T., Muelder, C. W., Faris, R., Felmlee, D., and Ma, K.-L. (2014). Visualization techniques for categorical analysis of social networks with multiple edge sets. *Social Networks*, 37:56 – 64.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

Cummings, J. N. and Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks*, 25(3):197 – 210.

David, E. and Jon, K. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA.

Davis, J. and Leinhardt, S. (1967). *The Structure of Positive Interpersonal Relations in Small Groups*. Darthmouth College.

Doran, D., Alhazmi, H., and Gokhale, S. (2013). Triads, transitivity, and social effects in user interactions on Facebook. pages 68–73.

Dorogovtsev, S. N. and Mendes, J. F. F. (2004). The shortest path to complex networks. *eprint arXiv:cond-mat/0404593*.

Everett, M. G. and Borgatti, S. P. (1999). The centrality of groups and classes. *The Journal of Mathematical Sociology*, 23(3):181–201.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks*, 19(2):157 – 191.

Fowler, J. H., Christakis, N. A., et al. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *British Medical Journal (BMJ)*, 337:a2338.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239.

Freeman, L. C., Romney, A. K., and Freeman, S. C. (1987). Cognitive Structure and Informant Accuracy. *American Anthropologist*, 89(2):310–325.

Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164.

Gondal, N. and McLean, P. D. (2013). What makes a network go round? exploring the structure of a strong component with exponential random graph models. *Social Networks*, 35(4):499 – 513.

Goodreau, S. M., Kitts, J. A., and Morris, M. (2009). Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography*, 46(1):103–125.

Granovetter, M. (1982). The strength of weak ties: A network theory revisited. In *Sociological Theory*, pages 105–130.

H R Bernard, P Killworth, D Kronenfeld, and Sailer, L. (1984). The Problem of Informant Accuracy: The Validity of Retrospective Data. *Annual Review of Anthropology*, 13(1):495–517.

Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459.

Hanneman, R. A. and M., R. (2005). *Introduction to Social Network Methods*. University of California, Riverside. http://faculty.ucr.edu/ hanneman/nettext/, USA.

Hubbell, C. H. (1965). An input-output approach to clique identification. *Sociometry*, 28(4):377–399.

Kamada, T. and Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7 – 15.

Katz, L. (1966). A new index derived from sociometric data analysis. *Psychometrika*, 18:39–43.

Kobourov, S. G. (2013). Force-directed drawing algorithms. page 383–408.

Kolaczyk, E. and Csárdi, G. (2014). *Statistical Analysis of Network Data with R*. Use R! Springer New York.

Kossinets, G. (2003). Effects of missing data in social networks. *eprint arXiv:cond-mat/0306335*.

Lusher, D., Koskinen, J., and Robins, G., editors (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press, Cambridge.

MVF (2015). Mamidipudi venkatarangaiya foundation (MVF). Accessed on 06 november 2015: http://mvfindia.in.

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, 67(2):026126.

NUEPA (2015). National university of educational planning and administration (NUEPA), indian unified district information system for education (UDISE). Accessed on 06 november 2015: http://www.dise.in.

Oliver Bastert, C. M. (2003). Layered drawings of digraphs. In Kaufmann, M. and Wagner, D., editors, *Drawing Graphs: Methods and Models*, Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.

Preciado, P., Snijders, T. A., Burk, W. J., Stattin, H., and Kerr, M. (2012). Does proximity matter? distance dependence of adolescent friendships. *Social Networks*, 34(1):18 – 31. Capturing Context: Integrating Spatial and Social Network Analyses.

Reingold, E. M. and Tilford, J. S. (1981). Tidier drawings of trees. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Software Engineering*, 7(2):223–228.

Ripley, R., Boitmanis, K., and Snijders, T. A. (2013). *RSiena: Siena - Simulation Investigation for Empirical Network Analysis*. R package version 1.1-232.

Robins, G., Elliott, P., and Pattison, P. (2001). Network models for social selection processes. *Social Networks*, 23(1):1 – 30.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173 – 191. Special Section: Advances in Exponential Random Graph (p*) Models.

Rodan, S. (2011). Choosing the ' ' parameter when using the bonacich power measure. *Journal of Social Structure*, 12.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.

Smith, J. A. and Moody, J. (2013). Structural effects of network sampling coverage i: Nodes missing at random. *Social Networks*, 35(4):652 – 668.

Smith, J. M., Halgin, D. S., Kidwell-Lopez, V., Labianca, G., Brass, D. J., and Borgatti, S. P. (2014). Power in politically charged networks. *Social Networks*, 36:162 – 176. Special Issue on Political Networks.

Snijders, T. A. and Koskinen, J. H. (2011). An introduction to stochastic actor oriented models. Presented at the Statistical Models for Social Networks, May.

Snijders, T. A., van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44 – 60. Dynamics of Social Networks.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.

Sugiyama, K., Tagawa, S., and Toda, M. (1981). Methods for Visual Understanding of Hierarchical System Structures. *IEEE Transactions on Systems, Man, and Cybernetics*, 11(2):109–125.

VAEIS (2015). Research project website: Value affirmation exercise in indian schools. (VAEIS). Accessed on 09 november 2015: http://dl.ucd.ie/˜pablo/-valuesaffirmation.

Vu, D., Pattison, P., and Robins, G. (2015). Relational event models for social learning in {MOOCs}. *Social Networks*, 43:121 – 135.

Wang, P., Robins, G., Pattison, P., and Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks*, 35(1):96 – 115.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press.

Wilkinson, L., Wills, D., Rope, D., Norton, A., and Dubbs, R. (2005). *The Grammar of Graphics*. Statistics and Computing. Springer New York.

Zhang, M. (2010). Social network analysis: History, concepts, and research. In Furht, B., editor, *Handbook of Social Network Technologies*, pages 3–21. Springer.