Children in time: community organization in social and ecological systems

Ferenc Jordán¹,²
¹Collegium Budapest, Institute for Advanced Study, Szentiáromság utca 2, 1014, Budapest, Hungary
²HAS Animal Ecology Research Group, Hungarian Natural History Museum, Budapest, Hungary

The hierarchical organization of many ecological and social systems calls for a better understanding of part-to-whole relationships. Network analysis provides a tool for this and it is routinely used for modelling interaction systems. I present a longitudinal social network analysis of a classroom focusing on properties ecologists would be most interested in. Analyses of ecological and social networks share many methodologies, and with many problems cross-relevant, I discuss the possibilities of cross-disciplinary thinking. I quantify the structural balance, the core–periphery organization, small-world character, the KeyPlayer-nestedness and the invalidity of this human community, over time, in the social network setting and look for ecological parallels.

Keywords: Food web, network analysis, phenology, social network.

Introduction

Both natural and social systems are immensely complicated: if we wish to understand them we need to choose sensible models for their study. Network representations of various systems may help understand the ‘part’ to ‘whole’ relationships: how do parts form the whole and how does the whole influences its parts? Answering these questions is non-trivial as distinct mechanisms operate at these two levels. The parts are more likely forming a population of similar entities, selected through evolution; they are best described by statistical tools. The whole, on the other hand, generally defies statistical descriptions, is more likely unique and is being transformed instead of selected for. Still, studying different levels of hierarchical organization in parallel is very important, because hierarchy is one of the main characteristics of life, in both nature and society.

To look at the parts in the context of the whole is not new. Gestalt theory and the various field concepts in different areas of science¹,² are the precursors of many recent studies on complex systems³. All of these approaches emphasize the fact that interactions between parts are not strictly localized: they spread indirectly but typically not to infinity, i.e. they have some effective range. Contrary to the old (soft) version of holistic thinking – simply stating that everything is connected to everything else – the current (harder) holistic studies provide quantitative results on what is connected to what and to what extent. This latter version can be predictive, basically testable and useful for science and society.

The well-established methodology of network analysis (cf. graph theory) is a simple and sometimes powerful (but sometimes irrelevant and inadequate) tool for quantifying all of the above. It is mostly useful in systems where interactions are of key importance. In ecological (e.g. a forest) and social (e.g. a classroom) systems, interaction is what makes the system a system, what makes a set of parts form a community, and what to study if the system is to be understood. In contrast, in certain systems, interaction is not of key importance: even if water molecules are connected to neighbours, a network model is not really adequate for modelling water, and although many small parts make a car, a network model of screws and parts may not help in better understanding the behaviour of a car. In fact, many ecologists believe that interactions between parts are even more important in ecosystems than the parts themselves⁴-⁶. Conventional wisdom (e.g. the enemy of my enemy is my friend; birds of a feather flock together) also reinforces the evidence that interactions are of high importance in social groups.

Because of the basic similarity of social and ecological systems, the similarity of the methods for their analysis (e.g. role colouration), and the similar problems they raise (transitivity of indirect effects, role of key positions, effects of system-level structure on functioning), there is growing interest in comparing them to each other⁸. Even if the particular components of the respective networks are different (human beings vs. species), important similarities may still exist given the importance of interactions in both models⁹,¹⁰. In both these systems, effects spread indirectly but not to infinity and they change while being transmitted (not like information on the internet: the message is the same whether you read your e-mail in Anchorage or Brisbane).

There are many current interdisciplinary studies on various networks¹¹, with the general aim to demonstrate the surprisingly similar properties of various systems, despite containing quite different parts. Note that while
comparing two static networks is highly sensitive to methodology, especially on network construction, similarly described networks are compared to each other the differences between their properties give more reliable information, since methodological noise is extinct. One way for such a comparative approach is to study time-series of networks, focusing on change, rather than static description. With this in mind, I present the phenology of the studied networks and try to compare them with similar ecological information (even if time series of ecological networks is incredibly rare).

Because of the shared problems, it is useful, or at least thought-provoking to study ecological questions in a social network context. By network, I mean a set of nodes (representing individual children) and a set of links between pairs of them. A link represents a relation between a pair of children: it can be either positive or negative (signed), it is directed from one towards the other (directed) and it is either weak or strong (weighted; although I consider only the strong ones). Networks have various properties providing potentially useful information for better understanding of their behaviour (group dynamics).

I will analyse the data of a longitudinal social network analysis survey and study several characteristics of the communities being highly interesting and central in current community ecology: (1) the role and identity of key positions in the network (cf. keystone species), (2) the general properties of the whole network, for example how small-world-like or how centralized it is (cf. ecosystem structure, organization), (3) the positions where new nodes can enter and nodes leave from (cf. invadability and extinction risk), (4) the pattern of positive and negative links and its effect on stability (cf. conflict and cooperation), (5) the ratio of nodes in the core and in the shell (cf. core and satellite species hypothesis), and (6) the stability of various network metrics in time series (essentially lacking in ecology).

It is very important to note that comparative network analysis has had tremendous problems, despite encouraging advancements\textsuperscript{12}. The typically taken static attitude is especially sensitive to methodological problems and may produce artefacts. However, as most of the dynamical studies are simulations, descriptive network time series are very rare. So, instead of presenting a very detailed, deep analysis, I only aim to sketch some of the most interesting and perspectival aspects of how to compare social and ecological community organization. I wish to suggest that future possibilities are endless and are worthy efforts\textsuperscript{8}.

Data
The data used in this study is a subset of a much larger dataset provided by a longitudinal social network analysis survey made in Hungary. Seven classes in three randomly chosen secondary schools were visited ten times during a three-year period (roughly evenly distributed). Children were asked to fill in a short, anonymous questionnaire asking for all their positive and negative relationships with classmates, as well as qualifying those relationships as either strong and weak. Questions were very simple in order to make interdisciplinary comparisons possible\textsuperscript{13}. In the present analysis, I have chosen one of the most complete datasets. I started monitoring the class half a year after they came together (January 2004, aged 11–12) and finished the survey before the final exams (February 2007), with eight more steps in between (I have altogether ten networks). Here I focus only on structural variables of students, not on their attributes (although I have data on their gender and their personal feeling in the class). Absent students were typically asked to fill in the questionnaire later, but sometimes their previous answers were incorporated into the series. Altogether four students arrived and two left the class during the survey.

Methods
I calculated several network indices taken from the toolbox of social network analysis (SNA; see ref. 14), partly used also in ecological network analysis (ENA). Some indices characterize the network as a whole, while others describe the network position of single nodes or nodal sets.

Network construction
Many methodological problems of network construction may be reduced in longitudinal surveys if the data are consistent. Similarly, biases in static databases may not influence trends or if they do, only to a lesser extent.

In this network the nodes were the children (defining nodes is less evident in food webs). Links were the relationships reported in the questionnaires. Weighted links were not considered per se but weak relationships were used as a filter for reliability. The sign of the links (either positive or negative) was used for constructing signed graphs (Figure 1 a and b) show positive and negative links respectively, Figure 1 c shows the total signed network and Figure 1 d shows the total unsigned network, where there is no difference between positive and negative links). While the direction of links is basically important, I do not consider this in the present analysis (it does not matter whether B likes A if A likes B). I study a time series of 10 signed, binary, symmetrical social networks (Figure 2).

Network analysis
In a graph, a loop is defined as a series of links with identical starting and end point (for example, the A–B, B–C
and C–A links form a loop of length three). In a signed graph, each link has a sign (either positive or negative) and each loop also has a derived sign: it is positive if the number of negative links is even, otherwise the loop sign is negative. Each dyad in the network (pair of nodes) was characterized by a single signed link: it was + for +/+ and +/0, it was − for −/− and −/0; in the ambiguous case of +/−, the overall sign was taken as negative although the ratio of such antagonistic relationships was very low. Note that each dyad could be characterized by two links in a directed network (also eliminating the problem of antagonistic links). According to a well-established theory of structural balance, a graph is structurally stable if all of its loops are positive. Harary also suggested a hypothesis that a network tends to change towards balance. I was interested in whether this theory is correct for these classes, but because in large networks such as the one described here it is statistically unlikely to have only positive loops, I used a weighted, non-binary version of structural balance. Here, non-binary means that I quantified the ratio of negative loops to total loops, i.e. structural balance had a value between 0 and 1. Weighting means that the contribution of each loop to the total loop set considered loop length: long loops mattered more, i.e. loop sign was weighted by length. Note that this kind of weighting can be done also in unweighted (binary) networks. I have considered loops with a maximal length of seven links. Thus, if \( L \) is loop length (\( L_{\text{max}} = 7 \)) and \( C^+ \) is the number of positive loops of length \( L \), then \( TS^+ = L \times C^+ \) measures the total number of links contributing to negative loops of length \( L \). Similarly for \( TS^- \) and TS. Finally, \( TS^-/TS \) is a weighted, non-binary measure of structural imbalance. According to Harary, this value is expected to decrease over time.

Based on a recent study, we are able to determine the centrality of nodal sets, not only single nodes. This is important, because it is possible, albeit rare that the most central node does not belong to the two most central ones, i.e. KeyPlayer sets are unnested. I have used the distance-based reachability (\( R_d \)) measure for determining the central sets of nodes. I quantified how nested were KeyPlayer sets in the negative and in the positive sub-graph, and how did they change over time.

The clustering coefficient (CL) of a graph node quantifies how densely its direct neighbours are connected to each other. In other words, it is the number of links between its neighbours divided by the theoretically maximum number of links between them. It can be averaged over the whole graph if one wants to provide a global network measure. Also, if given for the whole...
graph, it can be weighted by the size of neighbourhood of each node; I used this ‘weighted overall’ version\textsuperscript{22}.

The average path length (\(d\)) of a graph is defined as the mean distance between pairs of nodes. For nodes \(i\) and \(j\), their distance is defined as the length of the shortest path between them (length meaning the number of edges connecting the two nodes). The average characterizes the whole network.

In regular graphs (where each node has exactly \(n\) neighbours, like on a lattice), both CL and average distance (\(d\)) are high. In random networks, both measures are low. Small-world networks are characterized as having mixed properties of the two, i.e. relatively high clustering but relatively low average distance. Thus, a high CL/\(d\) ratio (denoted by SW) indicates that the given network is more ‘small world-like’\textsuperscript{23}.

Closeness centrality (CC) is one of the standard centrality indices in SNA. It is based on the summed distances of a node from other nodes in the network. Values were divided by the summed distances for the whole class, in order to make them comparable. Based on the ranking of nodes, I defined arbitrary thresholds for centrality groups of nodes. There were a core group of highest centrality, a periphery group of lowest centrality and three more groups in between. Thus, I allowed empty cores\textsuperscript{24} and decided to characterize the core-periphery structure by a gradient classified into five classes. When CC is calculated\textsuperscript{22}, \(+/+, +/0\) and \(+/-\) interactions are equivalent, while \(-/-, -/0\) and \(0/0\) relationships are the same as well, in other words, a link is permeable if it has at least one positive sign. It could be discussed whether this is logical for \(+/-\) relationships, but these are very sparse in our social networks, so the possible critique is also of minimal interest. However, if I do believe that \(+/-\) interactions are also able to transfer communication. The impermeable nature of \(-/-\) and \(-/0\) (and, naturally, \(0/0\)) relationships seems to be realistical, since these negative relationships seem to mean more the lack of communication than negative effects expressed. Based on the CC rank, I study the number and identity of students belonging to the core and to the periphery, according to different threshold levels (0.04, 0.035, 0.03 and 0.25). Some students never belong to the core, as their values are consistently lower than 0.025. I was also interested in how stable is the layer-assignment of students in time. I performed the same analysis for webs containing also the negative links but set unsigned.

Betweenness centrality (BC) is another standard centrality measure in SNA. It measures the probability that a given node is incident to shortest path between any pair
of other nodes. I have ranked nodes based on BC and looked for the rank values of individuals arriving to the class or leaving from the class. In the presented class, there was only four arriving and two leaving children. Rank positions of focal nodes were divided by the number of nodes, this is why this measure is called the relative importance rank (RIR). For example, if a newcomer is the most central child in the classroom, its RIR equals \(1/N\), if the classroom has \(N\) students (and if (s)he is the last one, RIR will equal 1). Since it does matter whether someone is the first out of 10 or out of 20, this measure is useful by making a difference between 1/10 and 1/20. In many networks, BC seems to be quite a relevant measure of centrality.

Results and discussion

Even as the ratio of negative and positive links remained roughly constant through time, the patterning of negative links has changed towards a decreasing ratio of negative loops (Figure 3a). In this respect, this classroom showed a behaviour supporting the hypothesized tendency towards structural balance. Regarding ecological networks as signed graphs is very poorly documented in ecological literature and ecologists simply lack data for this test. However signs should be taken more seriously in ecological networks, particularly given the findings that positive, negative and unsigned total networks show quite different properties and dynamics. It is also noted that the ratio of negative and positive links remained roughly constant, i.e. it is only their patterning that changed. If this tendency is general, it supports a very important message: it is not the level of conflict what needs to be understood but how it is combined with cooperation.

Determining central nodal sets and their nestedness has been proposed as a novel way to estimate the efficiency of conservation biology: if the most important node is one of the two most important ones, and so on, we may believe that setting priorities is a linear optimization problem. However, if this is not the case, an untested pattern of KeyPlayer sets suggests that efficiency is limited by highly complicated optimal patterns. In the classrooms, I checked KeyPlayer nestedness based on the \(R_d\) measure, for both the positive and the negative subgraph. Nestedness was simply characterized by the number of gaps in the first three KeyPlayer sets (a gap appears if a node in a smaller set is not a member of a larger set). I have found that in positive networks nestedness increases (gaps are fewer), while in negative networks it decreases (Figure 3b). Thus, positive networks are more controllable.

There is another marked difference between positive and negative networks in case of the small-world property (SW). The positive network is more small world-like, while the negative one does not change and has a much lower SW-value (Figure 3d). I analysed the total unsigned version as well and found that SW increased even more than for the positive network (Figure 3c). The strength of a tendency like this is a measure of organisation or health, in a sense. In the ecological literature it has been suggested that food webs may be SW networks but the evidence is currently poor (also the interpretation of its possible biological meaning).

In ecology, it is an old yet poorly studied question whether communities contain core and satellite species, i.e. relatively more and less important species for community organization. Concepts and measures are problematic, of course. Here, I decided to follow a very soft approach. I define five thresholds for separating the core and the periphery and analyse the differences in node affiliation, depending on the threshold. Using node centrality (CC) makes it possible to see empty and multiple cores appearing. Figure 3e shows the ratio of nodes entering the core at different threshold values: the core is always the bottom category. Figure 3f shows that the number of nodes in the lowest category (nodes always belonging to the core) is decreasing, while the other extreme, nodes always belonging to the periphery is increasing. Beyond the pure number of nodes belonging to the core, the other question is the identity of nodes. One can see a surprisingly large turnover of nodes in different layers. If negative links are also considered in ecological networks, the network becomes topologically more homogenous (see a case study). I was also interested in the centrality rank of nodes representing children newly coming to as well as just leaving the class. Centrality was measured by BC. RIRs have been calculated in order to keep these comparable. Figure 3g shows the RIR of four arriving and then two leaving individuals, in both the positive and negative networks. Here the central positions in the positive network are hidden: both arrivals and leavings happen only in less central positions. This finding suggests that it is easier to enter the centre of the negative network. The relationship between topological position and extinction and the effect of topology on invasion success are hot topics in current ecology. In these fields, the basic question is the same: how network position influences success in group dynamics. Also, RIR values characterize the invadability of a community: how to play a large role in the community as a newcomer (cf. Genghis Khan species, naiveté, home-field advantage). Even if social network data are very sparse (four new children in three years), there is much less data in ecology on invading food webs (interaction networks, but see ref. 33).

Conclusions

Both ecological and social networks raise similar questions, similar methodical problems, similar data types yet with strikingly different agents, i.e. people and species,
Figure 3. Eight charts illustrating the phenology of the presented classroom. 

**a.** the weighted negative to total ratio of loops (TS-TS). **b.** the number of gaps in the KeyPlayer sets in the positive (full triangle, broken line) and the negative (empty square, solid line) subgraphs. **c.** the small-world character of the total unsigned network (SW). **d.** the small-world character of the positive (empty square, broken line) and the negative (full triangle, solid line) subgraphs. **e.** the number of nodes belonging to different layers (blue is the core, purple is the periphery). **f.** the number of nodes in the core (empty circle, solid line) and in the periphery (full triangle, broken line). **g.** the RIR value of arriving (black and white columns) and leaving (striped and white columns) individuals in the positive (black and striped bars) and in the negative (white bars) subgraphs (at various time steps). X-axis is always time in ten steps (but in G).
characterize social and ecological communities. Although exchanging methodological tools had been recognized early, the lack of high-quality data still limits serious comparative studies. However, the goals are clear, for example, to better understand the possibly shared laws governing social and ecological communities. There is currently a strong interest in applying network analysis for studying animal groups and also for comparing human and animal groups (social wasps and children). It has also been suggested that applications may be looked for in human security issues. To be linked means both possibilities and constraints, and to better understand both of them may really contribute to predicting and even partly controlling the behaviour of these systems.


ACKNOWLEDGEMENTS. I thank the anonymous students and teachers for their cooperation. I also thank Marco Scotti for help in the analysis of signed graphs, Zsófia Benedek for preparing the core and periphery analysis and Zsuzsanna Benedek for technical help. I thank the Society in Science: The Branco Weiss Fellowship, ETH Zürich, Switzerland for financial support.