The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

How to Avoid Weight Inconsistencies in Social Networks – an Entropy-Driven Approach

Andreas Dellnitz*
Department of Quantitative Methods,
FernUniversität in Hagen, Germany

Elmar Reucher
Department of Business Administration,
Private Hochschule für Wirtschaft
und Technik, Germany

Wilhelm Rödder
Department of Operations Research,
FernUniversität in Hagen, Germany

Abstract

Social network analysis is a popular field of research and most commonly supported by graph theory-based methods. But there is an up-and-coming approach that has spilled over from the area of artificial intelligence – so-called entropy-driven social network analysis. This approach uses probabilistic conditionals to express relationships between actors or groups of actors rather than merely edges or arrows. The new method allows for calculating all actors' importance in the net. However, if an analyst imprudently assigns probabilities to conditionals in the social network, the whole structure can be inconsistent, and hence the above-mentioned indices then are meaningless. Therefore, we propose a new interactive algorithm that helps the analyst to detect and revise such inconsistencies. The applicability of the algorithm will be exemplified by a mid-sized family network.

Introduction

Studying Social Networks (SN) supported by quantitative methods is becoming increasingly important and is being continuously developed due to the strong demand for social media products such as Facebook or Twitter. However, theory of social fabrics or networks is much older than those platforms. In the 1930s, the sociologist Jakob Moreno laid the foundation for graph theory-based analyses of social fabrics; for the first time, he presented 1934 a graphical image of a set of actors (nodes) and their relationships (edges) – called sociogram, cf. (Moreno 1934). Since then, graph theory-based toolboxes have been continuously enriched by indices, e.g. to assess the importance of actors or relationships, and algorithms, e.g. to reveal the vulnerability of paths between actors or to detect groups or cliques. An overwhelming portrayal of the mathematical concepts can be found in (Newman 2012).

A new up-and-coming framework for modeling and analyzing social networks follows the above-mentioned classical graphical representation, but replaces the semantics of edges by conditionals – if-then. This means: if an actor i has, for example, a message, then actor j also does when both are connected by an arrow, refer to (Rödder, Brenner, and Kulmann 2014). Here, the authors show that under this conditional-logical framework one can determine diffusion and reception potential of each actor in the SN. For this purpose, they determine a probability distribution Q on all possible states of the network, using the principle of maximum entropy or minimum cross-entropy; for more details on this see (Csiszár 1975; Shore and Johnson 1980; Kern-Isberner 1998). A concept for modeling such conditional-logical frameworks are markov nets, which can be realized in the expert system shell SPIRIT (SPIRIT 2011).

However, one shortcoming of the new findings is that the authors restrict their analysis to certain conditionals: all conditionals have the same weight, i.e., probability 1. In (Rödder, Kulmann, and Dellnitz 2016), consequently, the authors go one step further and relax this restriction. Here, they show that if the probabilities of all conditionals are equal, one can always determine a Q which allows for studying the aforementioned indices. For different transfer probabilities, such a Q might not exist. Therefore, generating consistent transfer probabilities is subject of this paper. In order to reach this goal, first, we make use of the fact that for a totally unconnected social network, we can always determine a Q. Next, we iteratively add new conditionals to the SN, fathoming its upper and lower probability limits beforehand to further guarantee the existence of Q. If the desired probability does not fall into these limits – meaning that the probability is not compatible with currently collected net data –, the respective probability will be revised.

The remainder of the paper is organized as follows: The next section introduces the concept of entropy-based knowledge processing in a probabilistic conditional logical framework. Then we pave the way to consistently weighted SN and present the new algorithm. Next, we apply the algorithm to a mid-sized family network. Finally, we conclude and point to future research.

*Correspondence to andreas.dellnitz@fernuni-hagen.de

Syntax, Semantics and Usefulness of MinREnt-Nets

The following definitions and statements originate from (Rödder, Kulmann, and Dellnitz 2016); consequently and for the sake of transparency, we employ most of their notation to quickly catch up our idea of generating consistent social networks.

Let $a_1, \ldots, a_n$ be n actors. Each actor a_i will be represented by a binary variable V_i, with $V_i = v_i$ and $v_i = i$ or $\bar{i}$. Where $V_i = i$ or $\bar{i}$ is the proposition that actor a_i has the message (i) or not ($\bar{i}$). Thus, a vector $v = (v_1, \ldots, v_n)$ contains possible states of this network. For pairs of actors (a_i, a_j), we call $\mathsf{Q}(V_j = j \mid V_i = i) = p_{ij}$ a probabilistic conditional; if actor i has the message or information, then also j does with transfer probability p_{ij}.

Some of the potential information flows and the respective transfer probabilities p_{ij} might be known empirically, conducting e.g. surveys, counting frequencies of message flows, etc. They reflect effectiveness of each link.

As indicated in the introduction, we are now seeking for a probability distribution Q on $\mathcal{V} = \{v\}$ that respects all observed transfer probabilities – TPs for short:

$$\begin{gathered} \mathsf{Q}(V_j = j \mid V_i = i) = p_{ij} \text{ for } (i,j) \in N \\ \text{with } N \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}. \end{gathered} \tag{1}$$

Such a distribution is called *net deployment*, refer to (Rödder et al. 2019). In general, however, the observed links and respective transfer probabilities do not fully determine the distribution Q. The expert system shell SPIRIT (SPIRIT 2011) picks up an outstanding net deployment Q from all possible distributions – called MaxEnt- or MinRent-deployment; see, for example, (Rödder, Reucher, and Kulmann 2006). The shell solves

$$\begin{aligned} &\mathsf{Q}^* = \arg\max H(\mathsf{Q}) = -\sum_v \mathsf{Q}(v) \log_2 \mathsf{Q}(v) \\ &\text{s.t. } \mathsf{Q}(V_j = j \mid V_i = i) = p_{ij},\ (i,j) \in N \end{aligned} \tag{2}$$

or equivalently

$$\begin{aligned} &\mathsf{Q}^* = \arg\min R(\mathsf{Q}, \mathsf{P}^0) = \sum_v \mathsf{Q}(v) \log_2 \frac{\mathsf{Q}(v)}{\mathsf{P}^0(v)} \\ &\text{s.t. } \mathsf{Q}(V_j = j \mid V_i = i) = p_{ij},\ (i,j) \in N \end{aligned} \tag{3}$$

(2) and (3), respectively, comply with all surveyed TPs p_{ij} and calculate the deployment Q^* with maximal entropy H and minimal relative-entropy R being P^0 the uniform distribution. What can we learn if an actor a_i sends the message: $V_i := i$? In our context, fixing a state in this way is called evidenciation; it can be performed by solving

$$\begin{aligned} &\mathsf{Q}^{**} = \arg\min R(\mathsf{Q}, \mathsf{Q}^*) = \sum_v \mathsf{Q}(v) \log_2 \frac{\mathsf{Q}(v)}{\mathsf{Q}^*(v)} \\ &\text{s.t. } \mathsf{Q}(V_i = i) = 1. \end{aligned} \tag{4}$$

Thus, evidenciation again means conditioning, now adapting Q to Q^* such that $\mathsf{Q}(V_i = i) = 1$. If Q^{**} is determined via (4), then one can calculate the probability $\mathsf{Q}^{**}(V_j = j)$ for each $j \neq i$. For an actor a_j, it is the probability to receive the message when actor a_i has dispatched it. However, this probability might be subject to indeterminacy due to poor net structure. Even though, one can determine an interval estimate – named indeterminacy interval (II) – in which it must fall, cf. (Reucher 2002) and (Rödder, Reucher, and Kulmann 2006):

1. Solve (3) with the only restriction $\mathsf{Q}(V_j = j \mid V_i = i) = \epsilon$, $\epsilon > 0$ being a sufficiently small number. Result $\bar{\mathsf{Q}}^{\ell*}$
2. Solve (3) for $\bar{\mathsf{Q}}^{\ell*}$ instead of P^0. Result $\bar{\bar{\mathsf{Q}}}^{\ell*}$
3. Calculate $\ell = \bar{\bar{\mathsf{Q}}}^{\ell*}(V_j = j \mid V_i = i)$
4. Solve (3) with the only restriction $\mathsf{Q}(V_j = j \mid V_i = i) = 1 - \epsilon$, $\epsilon > 0$ being a sufficiently small number. Result $\bar{\mathsf{Q}}^{u*}$
5. Solve (3) for $\bar{\mathsf{Q}}^{u*}$ instead of P^0. Result $\bar{\bar{\mathsf{Q}}}^{u*}$
6. Calculate $u = \bar{\bar{\mathsf{Q}}}^{u*}(V_j = j \mid V_i = i)$
7. $[\ell, u]$ then is the indeterminacy interval (II)

Algorithm 1: computing II for $V_j = j \mid V_i = i$

A net deployment Q is called useful if for all i the relation $0 < \mathsf{Q}(V_i = i) < 1$ holds, otherwise it is named useless. For a justification refer to (Rödder, Kulmann, and Dellnitz 2016). Simply speaking, in the case of $\mathsf{Q}(V_i = i) = 0$ or $\mathsf{Q}(V_i = \bar{i}) = 0$, the conditional probabilities $\mathsf{Q}(V_j = j \mid V_i = i)$ and $\mathsf{Q}(V_j = j \mid V_i = \bar{i})$ are degenerated. If any deployment becomes useless in Algorithm 1, the determination of $[\ell, u]$ is impossible. The reason for this problem is an inappropriate assignment of TPs p_{ij} in (2) and (3).

In the following section, we show how the calculation of indeterminacy intervals can support the analyst by correcting inappropriate TPs .

From Useless towards Useful Social Networks

As indicated earlier, modeling a conditional probabilistic social network – with inappropriate transfer probabilities p_{ij} – can lead to useless network structures. This kind of issue can have various reasons: measurement errors in data acquisition, errors in data documentation, errors in data interpretation, lack of objectivity, insufficient reliability and validity, undiscovered outliers, improper data transformations into probabilities; for a deeper discussion regarding the latter refer to (Borgatti and Halgin 2011).

Unfortunately, often one cannot avoid all errors; therefore, we propose an algorithm which enables the analyst to modify interactively useless social network data. Step-by-step, we implement each conditional and the corresponding transfer probability – perhaps altered after checking its reliability. To get more familiar with the underlying philosophy of our algorithm, first, the procedure will be formulated in narrative terms; more technical details will then be given in Algorithm 2. To start with: Let us assume that for some (a_i, a_j), with $i, j = 1, ..., n$, the sociologist has observed respective TPs p_{ij}. Applying the aforementioned semantics, we can write a rule – a conditional and its probability – for an arbitrary (a_{i_o}, a_{j_o}) as $\mathcal{R}^{i_o, j_o} = \{\mathsf{Q}(V_{j_o} = j_o | V_{i_o} = i_o) = p_{i_o j_o}\}$. Furthermore, the sociologist might attribute reliability factors to such TPs, thus implying rule priorities. Let $\mathcal{R}_1, \ldots, \mathcal{R}_l, \ldots, \mathcal{R}_L$, with $\mathcal{R}_1 \succeq \mathcal{R}_2, \ldots, \succeq \mathcal{R}_L$, be the sequence of rules and $p_1, \ldots, p_L$ respective probabilities. Hence, $\mathcal{R}_l$ is the rule with the l-th priority for the pair (a_{i_l}, a_{j_l}) and $p_l := p_{i_l j_l}$ its underlying transfer probability. Given such a priority list for L pairs of actors, we call $\mathcal{R}^{des} = (\mathcal{R}_1, \ldots, \mathcal{R}_L)$ the *desired* set of rules, with $\mathcal{R}_1 \succeq \mathcal{R}_2, \ldots, \succeq \mathcal{R}_L$. When solving (2) or (3) by applying $\mathcal{R}^{des}$ and when the result is a useful Q^*, then the algorithm should immediately stop.In the case of inconsistent rules, however, we initialize ($l = 0$) the iterative procedure with an empty set of rules $\mathcal{R}^{sn} := \emptyset$. Thus, the actual uniform distribution P^0 is the current net deployment. In the first iteration ($l = 1$), we check the consistency between the rule with the highest priority $\mathcal{R}_1$ and the still empty set $\mathcal{R}^{sn}$. Needless to say, consistency in this case is always ensured, hence let $\mathcal{R}^{sn} := \mathcal{R}_1$. The second iteration comprises the same steps as the first one, yet applying the second rule $\mathcal{R}_2$. If $\mathcal{R}_2$ is not in contradiction to $\mathcal{R}^{sn} = \mathcal{R}_1$, we can implement $\mathcal{R}_2$ with p_2; the situation now is given by $\mathcal{R}^{sn} := \mathcal{R}_1 \cup \mathcal{R}_2$. However, let us assume that we detect an inconsistency in the l-th iteration. This is the case when the desired probability p_l is not an element of the indeterminacy interval (II), so $p_l \notin [\ell_l; u_l]$. Consequently, p_l must be revised. It is most likely that the analyst selects the upper or lower bound of the interval. The described procedure will be repeated, until the remaining rules are stored in $\mathcal{R}^{sn}$.

Generally speaking: the social network will be successively enriched by consistent relationships until its full implementation is realized. After this narrative delineation, Algorithm 2 details all steps in pseudocode.

input: a useless set of rules with a given priority sequence
output: a useful social network
1. *initialize an unconnected* SN *with* n *actors;*
2. $\mathcal{R}^{des} := (\mathcal{R}_1, \ldots, \mathcal{R}_l, \ldots, \mathcal{R}_\mathrm{L})$;
3. $\mathcal{R}^{sn} := \emptyset$;
4. $l := 1$;
5. **while** $l \neq L + 1$ **do**
6. *solve* $\bar{\mathsf{Q}}^{\ell*} = \text{arg max } H(\mathsf{Q})$ *with*
7. $\mathsf{Q}(V_{j_l} = j_l | V_{i_l} = i_l) = \epsilon$, $\epsilon > 0$;
8. *solve* $\bar{\bar{\mathsf{Q}}}^{\ell*} = \text{arg min } R(\mathsf{Q}, \bar{\mathsf{Q}}^{\ell*})$ *with* $\mathsf{Q} \models \mathcal{R}^{sn}$;
9. *calculate* $\ell_l = \bar{\bar{\mathsf{Q}}}^{\ell*}(V_{j_l} = j_l | V_{i_l} = i_l)$;
10. *solve* $\bar{\mathsf{Q}}^{u*} = \text{arg max } H(\mathsf{Q})$ *with*
11. $\mathsf{Q}(V_{j_l} = j_l | V_{i_l} = i_l) = 1 - \epsilon$, $\epsilon > 0$;
12. *solve* $\bar{\bar{\mathsf{Q}}}^{u*} = \text{arg min } R(\mathsf{Q}, \bar{\mathsf{Q}}^{u*})$ *with* $\mathsf{Q} \models \mathcal{R}^{sn}$;
13. *calculate* $u_l = \bar{\bar{\mathsf{Q}}}^{u*}(V_{j_l} = j_l | V_{i_l} = i_l)$;
14. **if** $p_l \in [\ell_l, u_l]$ **then**
15. $\mathcal{R}^{sn} := \mathcal{R}^{sn} \cup \{V_{j_l} = j_l | V_{i_l} = i_l\, [p_l]\}$;
16. $l := l + 1$;
17. **else**
18. **if** *some* $\tilde{p}_l \in [\ell_l, u_l]$ *is acceptable*
19. $\mathcal{R}^{sn} := \mathcal{R}^{sn} \cup \{V_{j_l} = j_l | V_{i_l} = i_l\, [\tilde{p}_l]\}$;
20. $l := l + 1$;
21. **else**
22. *stop algorithm;*
23. **end**
24. **end**
25. **end**

Algorithm 2: Generating useful social networks

actor	person	acronym
a_1	author	EGO
a_2	father	DAD
a_3	brother	BRO
a_4	aunt 2	AUN2
a_5	uncle 2	UNC2
a_6	grandaunt	GRAN
a_7	male cousin 1	COM1
a_8	male cousin 3	COM3
a_9	husband of female cousin	HCOF
a_{10}	second (female) cousin	SCOF
a_{11}	mother	MOM
a_{12}	sister	SIS
a_{13}	aunt 1	AUN1
a_{14}	uncle 1	UNC1
a_{15}	grandma	GRMA
a_{16}	female cousin	COF
a_{17}	male cousin 2	COM2
a_{18}	husband of sister	HSIS
a_{19}	wife of male cousin 3	WCM3
a_{20}	second (male) cousin	SCOM

Table 1: Acronyms of family members.

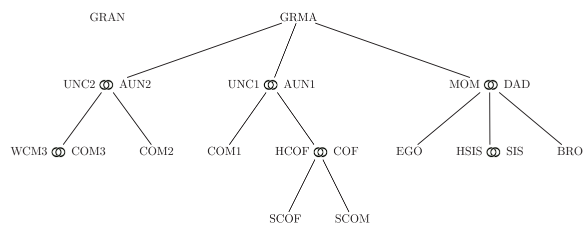

Figure 1: Pedigree.

If in line 18 the indeterminacy interval is inacceptable, the algorithm terminates. In this case, a revised priority list might be a way out of the deadlock.

Either way, the algorithm's success crucially depends on the willingness of the analyst to accept adjustments among troublesome transfer probabilities.

Information Flow within a Family Network

Subject of analysis

The following example, which contains 20 family members of different degrees of relation, originates from a bachelor thesis of the University of Hagen, Germany, cf. (Erlekotte 2013). For the actors see Table 1, being EGO the thesis's author. Figure 1 shows the corresponding family pedigree.

Table 2 depicts the information flows and priority order within the family network on forthcoming family celebrations and the respective transfer probabilities.

Interactively towards a useful family network

In this section, we apply the Algorithm 2 to the so far useless SN given in Table 2. We learn that the first 15 rules can be included without causing inconsistency, cf. Figure 2. However, if rule $\mathcal{R}_{16}$ with $p_{16} = 0.9$ would be added to $\mathcal{R}^{sn}$, the net deployment became useless. For the conditional UNC2 | AUN2, the indeterminacy interval yields $[\ell_{16}; u_{16}]$

l	relation l	p_l	l	relation l	p_l
1	UNC2 | COM2	[0.7]	19	GRMA | COF	[0.9]
2	HCOF | COF	[0.6]	20	BRO | MOM	[0.7]
3	COM2 | AUN2	[0.6]	21	MOM | AUN2	[0.8]
4	GRMA | GRAN	[0.9]	22	DAD | MOM	[0.9]
5	UNC1 | AUN1	[0.6]	23	GRMA | COM1	[0.9]
6	COM3 | WCM3	[0.6]	24	MOM | COM3	[0.6]
7	SIS | MOM	[0.7]	25	COF | COM1	[0.9]
8	EGO | AUN2	[0.6]	26	MOM | GRMA	[0.8]
9	SCOF | COF	[0.9]	27	GRMA | AUN1	[0.8]
10	AUN2 | COM2	[0.9]	28	GRMA | COM3	[0.9]
11	MOM | DAD	[0.9]	29	EGO | MOM	[0.8]
12	HSIS | SIS	[0.7]	30	MOM | AUN1	[0.6]
13	BRO | EGO	[0.7]	31	BRO | SIS	[0.7]
14	SCOM | COF	[0.9]	32	GRMA | MOM	[0.8]
15	SCOF | GRMA	[0.6]	33	COM1 | AUN1	[0.6]
16	UNC2 | AUN2	[0.9]	34	DAD | AUN2	[0.6]
17	HCOF | COM1	[0.6]	35	COM1 | COF	[0.9]
18	AUN1 | COF	[0.6]			

Table 2: Desired set of rules $\mathcal{R}^{des}$

= [0.40; 0.87]. This interval shows that the analyst should revise its perception about the transfer probability $p_{16} = 0.9$; hence, a value $\tilde{p}_{16}$ between 0.4 and 0.87 should be chosen to maintain the usefulness of the current net deployment. When refusing any adjustment, Algorithm 2 terminates. If the sociologist accepts the probability $\tilde{p}_{16} = 0.87$, we get a useful net deployment for $\mathcal{R}^{sn} := \{\mathcal{R}_1 \cup \cdots \cup \mathcal{R}_{15} \cup \{\mathsf{Q}(\text{UNC2} \mid \text{AUN2}) = 0.87\} \cup \mathcal{R}_{17} \cup \cdots \cup \mathcal{R}_{35}\}$, cf. Figure 2.

Conclusion and future work

When studying data, one has to deal with many measuring problems: e.g. too small and/or distorted samples, outliers, etc. Of course, these issues can also show up in social network analysis; and, most seriously, they can lead to inconsistent data when modeling social relations via probabilistic conditionals. The main goal of this paper is to eliminate possible inconsistencies. Therefore, we develop an interactive algorithm that helps the sociologist to revise inconsistencies. In order to run the algorithm, a prioritization of all probabilities is expected from the sociologist. How do priority orders influence the net structure? Answering this question might be an ambitious challenge for future research.

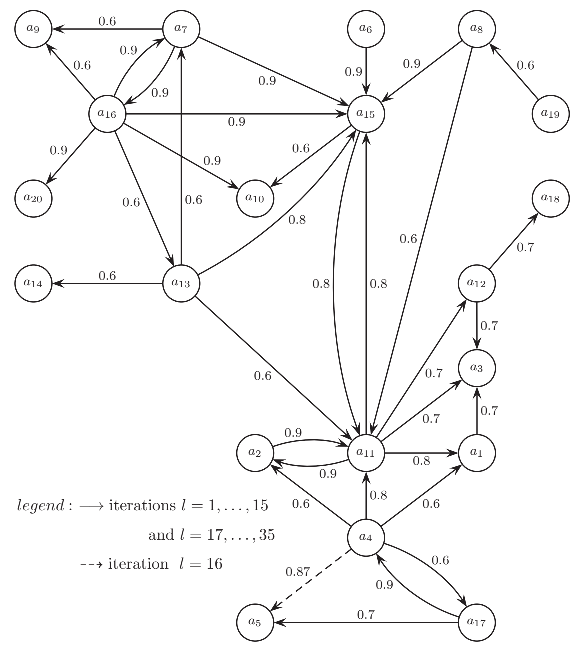

Figure 2: Useful family network.

References

Borgatti, S. P., and Halgin, D. S. 2011. Analyzing affiliation networks. In Scott, J., and Carrington, P., eds., *The SAGE Handbook of Social Network Analysis*. Sage Publications. 417–433.

Csiszár, I. 1975. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability* 3(1):148–158.

Erlekotte, H.-E. 2013. Modellierung sozialer Netzwerke in SPIRIT. *Bachelor thesis at FernUniversität in Hagen*.

Kern-Isberner, G. 1998. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence* 98(1-2):169–208.

Moreno, J. 1934. *Who Shall Survive: A New Approach to the Problem of Human Interrelations*. Washington, DC: Nervous and Mental Disease Publishing Co.

Newman, M. 2012. *Networks: An introduction*. Oxford University Press.

Reucher, E. 2002. *Modellbildung bei Unsicherheit und Ungewißheit in konditionalen Strukturen*. Berlin: Logos Verlag.

Rödder, W.; Dellnitz, A.; Gartner, I.; and Litzinger, S. 2019. Weight prediction on missing links in social networks: A cross-entropy-based approach. *Journal of Applied Logics – IfCoLog Journal of Logics and their Applications* 6:95–116.

Rödder, W.; Brenner, D.; and Kulmann, F. 2014. Entropy based evaluation of net structures deployed in social network analysis. *Expert Systems with Applications* 41(17):7968–7979.

Rödder, W.; Kulmann, F.; and Dellnitz, A. 2016. A new rationality in network analysis – status of actors in a conditional-logical framework –. In Beierle, C.; Brewka, G.; and Thimm, M., eds., *Computational Models of Rationality*, volume 20. College Publications. 348–364.

Rödder, W.; Reucher, E.; and Kulmann, F. 2006. Features of the expert-system-shell spirit. *Logic Journal of IGPL* 14(3):483–500.

Shore, J., and Johnson, R. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 26(1):26–37.

SPIRIT. 2011. Last accessed on 2019-10-22.