The Problem of Collecting Data on Mid-Sized Complete Networks (50 < n < 200) via Questionnaire: A Proposition for Using Subgroup-Based Name Generators

Grieser, Christopher

Erstveröffentlichung / Primary Publication
Arbeitspapier / working paper

The Problem of Collecting Data on Mid-Sized Complete Networks ($50 < n < 200$) via Questionnaire: A Proposition for Using Subgroup-Based Name Generators

---

Christopher Grieser
Technical University of Berlin
Berlin, Germany

Author Contact Information

- Mail: christopher.grieser@tu-berlin.de
- Address: Fraunhoferstraße 33-36, Sekretariatszeichen FH 9-1, 10587 Berlin, Germany
- Website: https://www.innovation.tu-berlin.de/v_menue/kollegiatinnen/3_ko-horte_042018_032021/grieser_christopher/parameter/en/

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The author would like to thank Leila Akremi, Robert Jungmann, Robert Panitz, and Nina Baur for helpful advice.

Abstract

Commonly, two approaches for collecting network data by means of a questionnaire are distinguished: rosters (complete name lists) and free recall (name generators) with subsequent merging of the ego-networks. However, both methods are reaching their limits when dealing with larger networks: rosters, on the one hand, increase in length with a larger network size, so that respondents either respond more unreliably due to fatigue, or they even abort answering the questionnaire all together. With free recall, on the other hand, weak ties and unpopular persons tend to be forgotten by the respondents, a problem that is also amplified with an increasing network size. In this paper, I want to propose an alternative method for collecting network data via questionnaire: the subgroup-based recall. With this method, it is possible to reliably collect data on mid-sized networks ($50 < n < 200$). When employing the subgroup-based recall, the network actors are divided into subgroups and a separate name generator is used for each group; the subgroups serve as cues without letting the questionnaire become too long. The use of the subgroup-based recall, however, leads to new methodological challenges, mainly the appropriate division of actors into subgroups. The discussion of subgroup determination shows that the method is best suited for collecting network data in organizational settings as they already provide detailed formal subgroups like for example departments. The article ends with specific recommendations for when to employ rosters, free recall and the subgroup-based recall.

Keywords: network measurement, network questionnaire, network data collection, respondent fatigue, questionnaire length, name generator, roster, mid-sized networks, intra-organizational networks

1. Introduction: The Problem of Collecting Data on Mid-Sized Networks with Questionnaires

Basically, there are two approaches for collecting network data with questionnaires: roster and free recall (e.g., Wasserman & Faust, 1994, p. 46). With rosters, the respondents are given a complete list of all other actors in the actor set and are asked to mark all the individuals with whom they have a specific tie. This not only requires that the researcher knows all members of the set prior to data collection, but also restricts the size of the network: when the set has more than a few dozen members, the roster will become too big and the questionnaire therefore too long (Herzog & Bachman, 1981; Rathod & LaBruna, 2005; Cape, 2010; Rolstad et al., 2011). With the free-recall method, the respondents are not presented a list of names but rather asked to name those persons with whom they have a specific tie. A list of names is thereby generated by the respondents and the specific instruments are hence known as name generators. However, forgetfulness of the respondents often leads to the omission of many ties, especially weaker ties (Ferligoj & Hlebec, 1999; Brewer, 2000; Marin, 2004). Yet again, this problem is amplified with a lager set of actors. Essentially, both methods are inadequate for gathering data on larger networks.

In this paper, I want to propose an alternative for collecting network data with questionnaires: the *subgroup-based recall*. The basic idea is to use a name generator for each subgroup of actors in the set that is to be investigated. The subgroups effectively function as cues of ties while avoiding an excessive length of the questionnaire. In practice, this method combines the strengths of rosters and free recall, and can reliably collect data on bigger networks – in my assessment, 200 actors are probably the upper limit. Although large networks are still a problem, this method enables the data collection of mid-sized networks (about $50 < n < 200$). While avoiding some limitations of rosters and free recall, the subgroup-based recall presents new challenges like the appropriate determination of subgroups and potentially more complex filtering in the questionnaire. Because of those challenges, the subgroup-based recall is probably most suited for data collection in more formal settings, e.g. the collection of intra-organizational network data.

First, I will present a short overview of the current methods for collecting network data via questionnaires and their shortcomings regarding larger networks (section 2). Then, I present the subgroup-based recall as a compromise and evaluate its questionnaire length compared with the roster (section 3). Afterwards, the main methodological challenge will be

discussed, namely the appropriate division of the actor set into subgroups (section 4). Based on the characteristics of each method, I will make recommendations which method should be used in which scenario (section 5). Finally, I will conclude by outlining potential further improvements for the subgroup-based recall (section 6). An exemplary implementation of the subgroup-based recall can be found in the appendix of the article.

2. Current Methods for Collecting Network Data via Questionnaires and their Shortcomings Concerning Mid-Sized Networks

While there is a variety of ways for collecting social network data – observation, small-world-techniques, archival records, or online data mining – questionnaires are still the most common method (Wasserman & Faust, 1994; Marsden, 1990, 2005; Baur, 2014). Their use can even become unavoidable when observations are not feasible and appropriate process-produced data is not available. The two basic approaches for collecting network data with questionnaires are the roster and the free recall method. With rosters, the respondents are presented a complete list of members in the set, while with free recall, respondents are asked to name all the people they have a specific tie with. While the exact phrasing of the question for identifying ties must be reflected for rosters and free recall, more attention has been paid to the phrasing of name generators (Bidart & Charbonneau, 2011) and well-known name generators are often referred to by the name of their creator, e.g. the Burt-Instrument (Burt, 1984).

Concerning the identification of ties, the two methods rely on two different psychological processes: with a roster, the respondents must *recognize* names, while with free recall, tie identification is based on respondents *recalling* names. The two methods also correspond to two different types of questions: rosters represent a series of closed questions (i.e. check-boxes), while the free recall method essentially constitutes one open question that asks for a (potentially long) list as an answer (Porst, 2014, pp. 53-70). Both methods also have certain requirements that must be fulfilled before the method can be employed properly. For rosters, prior knowledge of all actors belonging to the network is necessary to create the name list, while with free recall this is not an issue as unknown members of the network can be recognized via the name generators – new respondents are identified via snowball sampling

(Baur, 2014). However, with free recall, the ego-network-data must be properly merged to a complete network as for example done by Kirke (1996) in her study of youths in a district.

Both approaches have several advantages and disadvantages which tend to mirror each other. The main advantage of rosters is that the names presented in the roster serve as cues for the respondents, leading to less omissions of contacts, i.e. a higher completeness of edges. In contrast, the fact that respondents tend to forget weak ties and unpopular persons while answering name generators results in a higher omission of contacts and a lower completeness of edges (Ferligoj & Hlebec, 1999; Brewer, 2000; Marin, 2004). This is especially problematic as weak ties are considered to be highly relevant in social network analysis (Granovetter, 1973, 1983). The number of forgotten ties depends on many factors and varies significantly across various studies that compared name recognition and name recall – in some of the studies reviewed by Brewer (2000), the share of forgotten ties is even higher than 50%. Although there have been concerns regarding the accuracy of reported ties in general since the 1970s (e.g., Bernard et al., 1982; Killworth & Bernard, 1976), it is well established in cognitive psychology that recognizing items is far more reliable than recalling them (Groome, 2014).

Nevertheless, the problem of omitted ties is naturally only an issue when weak ties are relevant to the research question or the type of tie that is to be identified. When conducting a network analysis of close emotional bonds, for example, weak ties may not be theoretically relevant and the name generator may be perfectly adequate. Furthermore, there is potentially the problem that the merging of ego-network-data becomes problematic with a larger set of actors (e.g., because there are many people with the same first name), but this subject has also not been discussed in the current literature. However, online questionnaires which enable matching with a name list in the background and auto-complete names can mitigate that problem.

However, free recall also has a distinct advantage over the roster-method, namely the length of the questionnaire. A key difference between questionnaires for non-relational data and network questionnaires is that with the latter, the length of the questionnaire increases proportionally with the number of respondents. The problem of increasing questionnaire length is even amplified when more than one type of tie needs to be identified. Although longer questionnaires do not necessarily lead to higher drop-out rates, they do fatigue the

respondents and result in rushed, less attentive answers (e.g., Herzog & Bachman, 1981; Rathod & LaBruna, 2005; Cape, 2010; Rolstad et al., 2011). While drop-outs are easy to identify in the questionnaires, respondent fatigue is harder to recognize which makes it more difficult to assess. The monotony of long name lists compared to other questionnaires with more varied items further exacerbates the issue. A rule of thumb for rosters mentioned in the literature is that the network should not include more than 30 individuals (Baur, 2014, p. 952), although this is probably a conservative estimation.[1] With the free recall method, however, the questionnaire length always stays the same and only the number of responses given may increase, although this does necessarily correlate with the size of the network.[2]

One technique for reducing the length of name lists is to add a filter, asking whether the respondent even knows people from certain subgroups (e.g., departments of an organization) and only present names from that subgroup. Nevertheless, the use of filters has two drawbacks: First, respondents who know people in many subgroups such as the ideal-typical broker (Burt, 1992) are still confronted with a long list of names. Second, for respondents who know quite a few people, the filter can even increase the number of items presented when only two types of ties are collected. In contrast, name generators consist of just one open question per type of tie. While central actors do have to provide longer answers when faced with name generators, the advantage of name generators lies in the fact that actors with whom the respondents do not have a tie with never occur during the interview, greatly shortening the questionnaire length.[3]

Essentially, the choice between rosters and free recall entails a trade-off between a higher response quality and a lower omission of potentially relevant weak ties. Table 1 provides a summary of the characteristics of rosters and free recall. Although the decision between

[1] It could be assumed that central actors are more prone to aborting a questionnaire or answering less reliably than non-central actors because they have to name more contacts. In that case, rosters would be particularly inadequate for larger networks, as central actors are often especially relevant. However, one could also assume that central actors are more motivated to answer a network questionnaire because they are more curious concerning social relationships while peripheral actors feel demotivated to be confronted with their social isolation.

[2] Dunbar (1992), for example, has often been cited for the argument that the average maximum number of ties one person can maintain is around 150. While one can critique his methodology or the transferability of his result to specific networks, the basic assumption that there is an upper limit to the number of relationships an individual can maintain is quite reasonable.

[3] I am assuming that the name generators are just used for the construction of a complete network and not for a characterization of the ego-networks of the respondents. In the latter case, the name generators would be accompanied by a series of name interpreters, which naturally increase the questionnaire length for the free recall method.

both methods can be based on the research question – rosters are for example preferable when weak ties are essential for the research endeavor – they both reach their limit when facing bigger networks with more than 30-50 members. Surprisingly, the challenge to collect data on larger networks is not even mentioned in reviews on network measurement (e.g., Marsden, 1990, 2005). Therefore, I want to propose the subgroup-based recall as an alternative method which combines the strengths of both, rosters and free recall, while counterbalancing their weaknesses.

	Roster (Complete Name List)	**Free Recall (Name Generator)**
psychological process for tie recognition	recognition of names	recall of names
type of question(s)	a series of closed questions (checkboxes)	one open question
requirements	prior knowledge of all actors belonging to the network	proper merging of ego-networks to a complete network
advantages	the complete list of names serves as cues for weak ties and unpopular persons	the questionnaire becomes much shorter, resulting in less respondent fatigue
disadvantages	Longer questionnaire with a higher probability of drop-out and response quality	weak ties and unpopular persons tend to be forgotten

Table 1: Comparison of Roster and Free Recall

3. The Subgroup-Based Recall

The subgroup-based recall is essentially a hybrid of roster and free recall. The basic idea is fairly simple: the network is divided into multiple subgroups and one name generator is used for each subgroup. Although there is anecdotal evidence that methods similar to the subgroup-based name generator are already in use, there does not seem to be any formal description or instruction for it. The basic idea to use multiple name generators is already well known; with position generators for example, the questionnaire contains name generators for different types of relationships (Lin et al., 2001). With the subgroup-based recall, however, the type of tie is the same for each subgroup. Here, the name generator is practically repeated, and the subgroups serve as cues for the respondents which support the recall of their ties. It is well known in cognitive psychology that using cues increases the number

of items respondents can recall (Tulving & Pearlstone, 1966; Groome, 2014; Einstein & McDaniel, 2014).

Instead of asking „Who did you talk to in the last 3 months?“, the questionnaire would include the question „With whom from [subgroup X] did you talk to in the last 3 months?“; and that question is then repeated for every subgroup in the network. To account for the fact that the respondents' knowledge of subgroup-membership of others is rarely perfect, I propose to add a „leftover“ name generator for contacts who cannot be assigned to a subgroup listed before.[4] Afterwards, the resulting ego-network-data is merged into a complete network, like with the free recall method. Any techniques used for name generators, like autocompletion of names or matching with a background name list, can also be employed for the subgroup-based recall.

Compared with the roster, the subgroup-based recall avoids an excessive questionnaire length as there is only one open question per subgroup and respondent fatigue should therefore be lower. Compared with the free recall, the number of omitted ties should be lower because the subgroups serve as cues for the respondents. While the psychological process of recognition is more effective than cued recall, cued recall is still better than free recalls (Groome, 2014; Einstein & McDaniel, 2014). Essentially, the subgroup-based recall tries to achieve a balance in the trade-off between respondent fatigue and omitted ties. In my estimation, the subgroup-based recall should enable data collection for networks up to 200 members, but that number should be tested in further evaluations to properly assess the upper limit. While certainly not perfect – some weak ties will still be omitted – the usage of the subgroup-based recall results in better data when compared to rosters and free recall.

[4] As a side effect, the subgroup-based recall also results in some data on the (perceived) subgroup membership of the network actors. It could be that one person is always associated with a certain subgroup although that person officially belongs to another. Such a finding could point towards relevant brokering activities of that individual.

	Roster	Free Recall	Subgroup-Based Recall
few omissions of (weak) ties	**+ + +**	**− − −**	**+**
short question-naire length	**− − −**	**+ + +**	**+ +**

Table 2: Comparison of strength and weaknesses of roster, free recall, and subgroup-based recall

Concerning the practical implementation, it is possible to use a filter asking whether the respondent even knows people from certain subgroups to further reduce the number of questions posed in the questionnaire, similar to the roster. The use of such filters is mostly intended for questionnaires where information on multiple types of ties are to be collected. However, such filters are not without drawbacks. When using written questionnaires, complex filters should be avoided, as they tend to confuse respondents and therefore lower the response quality (Klöckner & Friedrichs, 2014, p. 679; Porst, 2014, pp. 155-160). For the identification of multiple ties, the subgroup-based recall is hence more suited for online questionnaires or personal interviews.

Having presented the basic characteristics of the subgroup-based recall, it is now possible to compare the length of the questionnaire for the subgroup-based recall and the roster. The total number of items needed when using the subgroup-based recall can be determined with the equation (1):

$$(1) \quad SR = 1 + (s+1)*t$$

Here, `SR` is the total number of items, `s` the number of subgroups named in the preceding filter question and `t` is the number of types of ties that are investigated. (The `1` added to the number of subgroups represents the leftover name generator.) If the subgroup-recall is used with 8 subgroups on average and if the average respondent has contacts in half of the subgroups, `s` would be 4. I further assume that `t` is 3 in most network questionnaires. Under those assumptions, `SR` should be 16 on average.

This number can be compared with the number of items for the roster. If the roster is used with a filter question for subgroups and the subgroups are of the same size, the total number of items `RO` for the roster can be calculated with equation (2):

$$RO = st + (n/st)*s\ *t \quad (2)$$

Here, n is the size of the network, st the total number of subgroups, s the number of subgroups named in the preceding filter, and t the number of type of ties. It is important to note that because open questions and closed questions differ concerning the effort for the respondents to answer them, SR and RO are not directly comparable. (As SR is mostly based on open questions which require more effort, the filter question in the beginning is counted as 1, while the filter question in RO is counted as st.) While t should be the same, regardless of the method (and therefore 3 for this estimation), it is not reasonable to assume that st and s are the same for the roster. It makes more sense to use a higher number of total subgroups, as more subgroup options can reduce the length of names list (n/st)*s more effectively. Assuming that the number is not too high, as any division of subgroup must still be known to the respondents, I will provisionally use the value 16 for st. I will assume that with a finer division, the number of subgroups named in the filter question will also be lower. For this estimation, I assume that it is three eighths instead of a half, so that s will be 6. Taking the supposed upper limit of n = 200 for the subgroup-based recall, the resulting value for RO is 241. And even with a smaller network size, like for example n = 100, RO is still 128.5.

As n does not appear in the equation (1) for SR (the number of items does not depend on the size of the network), its value is 16 for both cases. Again, SR and RO are not directly comparable, as the former represents open questions and the latter represents simple check-boxes. Nonetheless, it seems reasonable to assume that respondent fatigue will be lower for respondents who need to answer 16 open questions, while it is quite likely that respondent fatigue will occur when checking a list of 128 or even 185 items, even if they only demand yes or no responses.

The comparison of the total number of items becomes even more clear when taking into account that not every respondent is an "average" respondent. One can consider for example an ideal-typical broker (cp. Burt, 1992): Based on the assumption that a broker knows people in every subgroup because of their strategic choice of relationships, it must be assumed that s = st. When inserting s = st in the equations (1) and (2), SR would be 28 while RO would be 616 for n = 200, and 316 for n = 100. So even when the roster is considered to be more effective for the average respondent, the theoretical relevance of

brokers is an important argument in favor of the subgroup-based recall.[5] The described illustrative calculations of the total number of items are summarized in table 4.

	Roster (n = 100)	**Roster (n = 200)**	**Subgroub-Based Recall**
average respondent	128.5 (checkboxes)	241 (checkboxes)	16 (open questions)
ideal-typical broker	316 (checkboxes)	616 (checkboxes)	28 (open questions)

Table 4: Illustrative calculation comparing the total number of items for rosters and subgroup-based recall, both with filters for subgroups in which the respondent knows at least one person. n *signifies the number of actors in the complete network (the total number of items for the subgroup-based recall does not depend on* n*). The "average" person is assumed to have contacts in 50% of the subgroups for the subgroup-based recall and 37,5% for rosters (if rosters allow a higher subgroup number and therefore more effective filtering). The ideal-typical broker knows individuals in every subgroup. This calculation further assumes 3 types of ties that are to be investigated.*

4. The Main Methodological Challenge: The Appropriate Division of the Actor-Set into Subgroups

The main challenge of the subgroup-based recall is the appropriate division of actors into subgroups. Although seemingly easy, this division is a non-trivial issue and not reducible to the boundary specification problem in network analysis (Laumann et al., 1983). Basically, any subgroup-division must fulfill seven conditions:

- saliency,
- appropriate number of subgroups,
- homogenous group size,
- flexibility of the division,
- complete coverage,
- mutually exclusive group membership,
- and pragmatic determination.

[5] Even if brokers hves a higher motivation to answer a network questionnaire due to their interest in social relationships, it should be noted that SR for a broker requires less effort than SR for the average respondent as the (ideal-typical) broker will reduce the number of their informationally redundant contacts (Burt, 1992) Therefore, they would have to name less names per open question.

Saliency of subgroups means that the respondents are able to assign their contacts to subgroups. This knowledge of the subgroup-memberships of others mainly depends on the social setting of the network: In many companies or state agencies, for example, the affiliation of individuals to certain departments tends to be very clear („Tom works in HR, while Linda works in Controlling"). Meanwhile in scientific communities, for example, it is often not very clear to which sub-community an individual belongs when one is not part of a specific sub-community. As even with social settings where membership in subgroups is clear to most members, that knowledge is rarely perfect. As there are sometimes contacts which cannot be assigned to a subgroup by a respondent, the use of leftover name generators is recommended for all cases. These considerations make it clear that any subgroup division should be based empirically on the perceptions of the actors (the realist approach in Laumann et al., 1983) instead of artificial divisions by the researcher (the nominalist approach).

The next three issues are closely related. The *number of subgroups* is relevant, as too many subgroups result in an increase of the number of questions, leading to respondent fatigue and its associated negative effects, similar to rosters. Too few subgroups, and the benefit of using subgroups as cues is too small and the subgroup-based recall basically becomes the free recall method. Depending on the number of types of ties that will be investigated, 6-10 subgroups are probably reasonable, although that number requires empirical testing. Furthermore, the subgroups should be of *homogeneous size*, as for example one large subgroup containing 90% of the individuals and eight smaller subgroups for the remaining 10% will not be very useful for the identification of contacts in the big subgroup. Heterogenous size in general will tend to result in more tie omissions in the bigger groups compared to the smaller ones, distorting the resulting network. To achieve those two conditions, any existing subgroup division must be *flexible* enough to allow the merging or further division of any division at hand. If, for example, the marketing department of a firm is very large, it could be reasonable to divide it into print marketing and online. And in case there are too many subgroups, a department of sociology and a department of political science can be merged into a bigger subgroup „social scientific university departments". Thus, the exact determination of the subgroup division can require some creativity. A pretest should evaluate whether the determined division makes sense for the surveyed.

Another requirement for the subgroup division is that it should *completely cover the whole set of actors*. Should the subgroups not cover all actors, those actors will be missing in the final network which obviously distorts the subsequent network analysis. However, this requirement is not strict, as the leftover name generator will still be able to compensate the issue to a certain degree. The next condition is the *mutual exclusivity of group memberships*, meaning that the subgroups possess as little overlap as possible. Should the subgroups have a high degree of overlap, the actors belonging to multiple subgroups are far more likely to be named, again distorting the subsequent network analysis. A low degree of overlap may be tolerable when factored in the ensuing analysis, e.g. by considering that a high centrality of those actors might be a methodological artifact.

The last requirement, the *simple determination of the subgroup division*, is the most pragmatic requirement. The determination of subgroups should be achieved easily, as it is only preparatory work for the actual data collection. When a subgroup division can only be determined by elaborate interviews, e.g. multiple prior interviews, the use of the subgroup-based recall might simply demand too many resources. Yet again, this requirement makes the use of existing subgroup divisions far more reasonable than any artificial determination of subgroups.

Considering all the criteria for an appropriate subgroup division, it becomes clear that the subgroup-based recall is mostly suited for organizational settings like companies, state agencies, or NGOs as their formal structure usually offers a wide range of fine subgroup divisions, that cover the whole organizations with little overlap. Moreover, the membership to those subgroups is usually salient to most non-recent organization members. Information on the formals structure of an organization is also mostly easily available. The adequacy of organizational settings becomes even more clear when compared to other settings, e.g. scientific communities: The membership of others in scientific sub-communities is often not clear to outsiders, the sub-communities have neither similar sizes nor are they flexible enough to derive a set of six to ten sub-communities. As scientists often belong to multiple sub-communities, their membership is obviously also not mutually exclusive and as newcomers are often not associated with any sub-community, the sub-communities also do not cover the whole actor-set. And finally, the identification of sub-communities can be rather difficult if there are no scientific associations with appropriate subgroups.

Considering the presented issues for subgroup division, it is easy to explain why the method of simply employing multiple name generators is unsatisfactory: the name generators rarely refer to subgroups of similar size, they do not cover the whole network, and the subgroup membership is often not mutually exclusive. Therefore, the resulting data is too distorted to construct a valid complete network out of the ego-network-data. Subgroup determination based on participation in events or activities (Breiger, 1974) is also not very useful, as the temporary nature of most events will likely result in weak salience for the respondents. The steps and requirements of the subgroup-based recall are summarized in table 5. For illustrative purposes, an implementation of the subgroup-based recall can be found in the appendix of this article.

#0	**Check whether an appropriate subgroup division is possible** • saliency • appropriate subgroups-number • homogenous group size • flexibility of the division • complete coverage • exclusive group membership • simple determination
#1	**Division of actors into appropriate subgroups** • 6-10 subgroups of roughly similar size • in a pretest, it should be tested whether a certain subgroup division works for the respondents
#2	**Construction of the questionnaire with three main parts** a) filter question: in which subgroups does the respondent have contacts? b) name generator for each named subgroup c) leftover name generator for contacts that cannot be associated with one of the listed subgroups (→ *b) and c) are repeated for every type of tie)*
#3	**Merging of the resulting ego-networks into a complete network**

Table 5: Steps for Implementing the Subgroup-Based Recall

5. Recommendations on when to use Rosters, Free Recall, and the Subgroup-Based Recall

In light of the characteristics of the three methods, it is now possible to make recommendations on when to employ each method based on its strengths and weaknesses. The roster is still the best method for identifying weak ties and ties with unpopular persons. Thus, if

the network is not too big, and if weak ties are relevant for the research question, e.g. networks of information diffusion, the roster is still the best method. For mid-sized networks, the usefulness of rosters may vary, depending on the effectiveness of filters. When brokers are very relevant for the research questions, e.g. in manager networks, any filtering of subgroups in which the respondent has contacts becomes problematic, leading to an exceedingly long questionnaire for those brokers. But when brokers are practically irrelevant for the research questions, rosters with good filtering may become suitable even for mid-sized networks. When using the roster for mid-sized networks, potential respondent fatigue should be taken into account. This can be accomplished by comparing the frequency with which contacts are identified at the beginning of the name list with the frequency with which contacts are recognized at the end of the name list. When the probability to be recognized correlates with an early position in the name list, respondent fatigue has most likely occurred which unfortunately indicates a lower data quality.

The free recall method is best suited for situations when the omission of weak ties is tolerable, e.g. networks of emotional bonds. Free recall may also be suited for research questions which are more concerned with actions that are based on the subjective perception of a network rather than partially unconscious (network) conditions of action, because ties not recalled are simply not relevant for that type of question. In contrast, the free recall method is less suited when weak ties are relevant, like in Granovetter's (1973) classic study of information networks for job openings. As the free recall method results in the shortest questionnaire length, the method is most appropriate for bigger networks when rosters as well as subgroup-based recall will both reach their limits. It is useful to employ the free recall with auto-completion methods which greatly reduce the "cognitive work" of the respondents when naming contacts.

The third method which I proposed here, the subgroup-based recall, derives its strengths from combining the benefits of both aforementioned methods. Subgroup-based recalls are more effective in the identification of weak ties, and they are more effective with brokers, as the questionnaire length is still kept within limits for those respondents. Thus, when weak ties as well as brokers are relevant for the research question, it is advisable to employ the subgroup-based recall. The method is also beneficial in most cases where the other two have their shortcomings. The subgroup-based recall should be employed when there is a

larger network which is relevant to even if the actors are not directly aware of it. A good example for this would be communication networks. However, the subgroup-based recall is limited to social settings where the aforementioned conditions of appropriate subgroup-divisions can be achieved. Minor violations of those conditions may be tolerable, but if one or more conditions cannot be fulfilled at all, rosters or free recall may be preferable. However, it should not be underestimated that most modern social settings can be linked with organizations in one way or the other. Friendship networks among high school students, for example, can be linked with classes and teachers of the respective class. Using this angle, there are probably more settings where the subgroup-based recall is suitable than one would expect at first. For situations where all three methods are equally adequate or equally inadequate, e.g. a research question concerning a non-organizational setting where brokers as well as weak ties are relevant, one must weigh the advantages and disadvantages of each method.

6. Conclusion

The discussion of the two main methods for collecting network data via questionnaire, roster and free recall, showed that both methods reach their limits when it comes to networks with more than 50 actors – the former because of respondent fatigue, and the latter because of increasing omission of ties, especially weak ties. In this contribution, I have proposed the subgroup-based recall as a third alternative which combines the strengths of both methods while trying to avoid their drawbacks. The use of cues makes the recall of contacts more effective for the subgroup-based recall than with the free recall method, as insights from cognitive psychology show. The subgroup-based recall also leads to a vastly shorter questionnaire than with the roster, as an illustrative calculation of the total number of items has shown. If the main methodological challenge, the determination of an appropriate subgroup division, can be overcome, it should be possible to reliably collect data on mid-sized, interpersonal networks up to a size of about 200 individuals.

As the subgroup-based recall is still an untested concept, some specifics – a reasonable upper limit for the network size, or the adequate number of subgroups – remain open for empirical investigation. Furthermore, there is still room for refinements. One idea could be to think one step further regarding the combination of rosters and free recall. For example,

if only one or two subgroups are named by a respondent in the first filtering question, the questionnaire could still „fall back" to a traditional roster for that respondent, presenting only names from the named subgroup(s). The potential offered by online questionnaires is surely not exhausted yet. Additionally, while organizational settings are certainly most suited for the subgroup-based recall, other social settings come to mind where the adequacy of the method is still debatable: Is the subgroup-based recall useful for networks in college, where the belonging to subgroups like classes is often temporary (changing every semester) as well as overlapping? The use of the subgroup-based recall in different specific settings is thus an interesting question for further research.

References

Baur, N. (2014). Netzwerkdaten. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 941–958). Wiesbaden: Springer VS.

Bernard, H. R., Killworth, P. D., & Sailer, L. (1982). Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. *Social Science Research*, *11*(1), 30–66. https://doi.org/10.1016/0049-089X(82)90006-0

Bidart, C., & Charbonneau, J. (2011). How to Generate Personal Networks: Issues and Tools for a Sociological Perspective. *Field Methods*, *23*(3), 266–286. https://doi.org/10.1177/1525822X11408513

Breiger, R. L. (1974). The Duality of Persons and Groups. *Social Forces*, *53*(2), 181–190. https://doi.org/10.1093/sf/53.2.181

Brewer, D. D. (2000). Forgetting in the recall-based elicitation of personal and social networks. *Social Networks*, *22*(1), 29–43. https://doi.org/10.1016/S0378-8733(99)00017-9

Burt, R. S. (1984). Network items and the general social survey. *Social Networks*, *6*(4), 293–339. https://doi.org/10.1016/0378-8733(84)90007-8

Burt, R. S. (1992). The Social Structure of Competition. In N. Nohria & R. G. Eccles (Eds.), *Networks and Organizations: Structure, form, and action* (pp. 57–91). Boston: Harvard Business School Press.

Campbell, K. E., & Lee, B. A. (1991). Name generators in surveys of personal networks. *Social Networks*, *13*(3), 203–221. https://doi.org/10.1016/0378-8733(91)90006-F

Cape, P. (2010). Questionnaire length, fatigue effects and response quality revisited. *Survey Sampling International*, *111*.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. Journal of Human Evolution, 22(6), 469–493. https://doi.org/10.1016/0047-2484(92)90081-J

Einstein, G. O., & McDaniel, M. A. (2014). Prospective Memory and Aging: When It Becomes Difficult and What You Can Do About It. In T. Perfect & D. S. Lindsay, The SAGE Handbook of Applied Memory (pp. 37–58). Thousand Oaks, CA: SAGE. https://doi.org/10.4135/9781446294703.n3

Granovetter, M. (1973). The Strength of weak ties. *The American Journal of Sociology*, *78*(6), 1360–1380.

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, *1*, 201–233. https://doi.org/10.2307/202051

Groome, D. (2014). Long-Term Memory. In D. Groome (Ed.), *An introduction to cognitive psychology: processes and disorders* (Third edition, pp. 156–203). London; New York: Psychology Press, Taylor & Francis Group.

Herzog, A. R., & Bachman, J. G. (1981). Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly*, *45*(4), 549–559. https://doi.org/10.1086/268687

Killworth, P., & Bernard, H. R. (1976). Informant Accuracy in Social Network Data. *Human Organization*, *35*(3), 269–286.

Kirke, D. M. (1996). Collecting peer data and delineating peer networks in a complete network. *Social Networks*, *18*(4), 333–346. https://doi.org/10.1016/0378-8733(95)00280-4

Klöckner, J., & Friedrichs, J. (2014). Gesamtgestaltung des Fragebogens. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 675–685). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18939-0_49

Laumann, E. O., Marsden, P. V., & Prensky, D. (1983). The Boundary Specification Problem in Network Analysis. In R. S. Burt & M. Minor (Eds.), *Applied Network Analysis* (pp. 18–34). Beverly Hills, CA: SAGE Publications.

Lin, N., Fu, Y., & Hsung, R.-M. (2001). The Position Generator: Measurement Techniques for Investigations of Social Capital. In N. Lin, K. S. Cook, & R. S. Burt (Eds.), *Social capital: theory and research* (pp. 57–82). New York: Aldine de Gruyter.

Marin, A. (2004). Are respondents more likely to list alters with certain characteristics? *Social Networks*, *26*(4), 289–307. https://doi.org/10.1016/j.socnet.2004.06.001

Marsden, P. V. (1990). Network Data and Measurement. *Annual Review of Sociology*, *16*, 435–463.

Marsden, P. V. (2005). Recent Developments in Network Measurement. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 8–30). Cambridge; New York: Cambridge University Press.

Porst, R. (2014). *Fragebogen. Ein Arbeitsbuch*. Wiesbaden: Springer Fachmedien Wiesbaden. Retrieved from http://link.springer.com/10.1007/978-3-658-02118-4

Rathod, S., & LaBruna, A. (2005). Questionnaire length and fatigue effects – does size really matter? In *ESOMAR, Conference on Panel Research, Budapest*.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health*, *14*(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 381–391.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge, New York: Cambridge University Press.

Appendix: Implementation of the Subgroup-Based Recall

[introduction, other relevant questions]

Your Contacts at [Company X]
Now, we want to ask you some questions concerning your contacts at [Company X]. First of all, in which departments of [Company X] do you know anyone? Please check all departments in which you know at least one person by name.

[checklist of all subgroups of Company X]
(→ this serves as filter for the subsequent questions)

Personal Contacts at [Company X]
With whom from [department #1] did you talk during the last three months about personal affairs that are not related to your work?
Please name as many persons as you can think of, be it one person or several persons. If you cannot think of any person with whom you talk about personal affairs in this department, then you would leave the field empty. If you are not sure whether a certain employee works at this department, please name them anyway.

[open field]

Personal Contacts at [Company X]
With whom from [department #2] did you talk during the last three months about personal affairs that are not related to your work?

[open field]
(→ repeat this question for every subgroup of Company X, in which the respondent knows someone)

Personal Contacts at [Company X]
Are there any other persons who come to your mind, who you do not associate with any previously mentioned department?

[open field]