

ASSIGNMENT 2 – THE CLASS NETWORK DATA

OUR DATA: NETWORK_CLASS.SAV

The dataset *Network_class.dta* contains the answers of our class survey and is organized as follows:

The variable [pcode] is your personal code. This was the code that you may have filled out at the beginning of the survey to be able to identify yourself in the data. Not everybody made use of this option, so this variable has several missing values.

The second variable is a (randomized) identifier [id]. You can use it to match/merge (parts of your) SPSS data to the network data that you need to create.

The variables [k1], [k2], ... represent the extent to which a given participant knows a given person. These data can be used to create a Ucinet data set. The easiest way to do this is by transferring the data-matrix into Ucinet. Consult the Ucinet manual on how to do this.

The rest of the variables speak for themselves (have a look at the variable labels and the rest of this document). For clarity, we include screenshots of the actual questions as you saw them on your screen in this document.

Take care in dealing with the data. I have tried to get rid of mistakes, but there might be some leftover trouble in there – be careful! Messing around with data, both in terms of reading them in, converting them, transforming them, and analyzing them, is part of the assignment and will take time.

The actual assignment and the logistics surrounding it you can find on page 8 and further.

Enjoy!

Chris Snijders

AN OVERVIEW OF (MOST OF) THE SURVEY QUESTIONS

Welcome at the social network survey!

Please fill out your student number. We use this only to be able to see whether you participated, not for anything else. The student numbers will not appear in the dataset.

Student number (use 6 digits, so leave out the leading zero):

If you want to be able to identify yourself in the data, add a personal code here consisting of 4 digits and make sure to remember it yourself. If you do not want to be able to identify yourself in the data, leave this question blank.

Personal code (optional):

The survey takes some time to complete, but please complete it in a single run. Stay focused!
Percentage of TU/e students who completed the survey in a single run: more than 75%.

Good luck!

[Next >](#)

The first one is your student number (no longer visible in the data you received), the second is the variable [pcode].

Please have a look at the three Next buttons on the bottom of the page. All of them will direct you to the next page. To keep track of the way you feel about the survey, we would like you to use them to indicate your satisfaction or dissatisfaction with the survey.

Click the red button to indicate irritation or being dissatisfied with the task you have to do on the given page. Conversely, clicking the green button on the right means that you liked the questions you got. The gray button in the middle indicates neutrality.

Click one of the buttons to start the experiment.

[Next :-|](#) [Next :-|](#) [Next :-\)](#)

These data are in the variables [q_button?].

The [aboutyou*] variables (screenshot shows the questions in random order)

HINT: Google for “Opinion Leaders and Opinion Seekers: Two New Measurement Scales”. Journal of the academy of marketing sciences, vol 24, no. 2: 137-147.

Please indicate to what extent the following applies to you:

	Does not apply at all				Applies completely
You actively attack problems.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Whenever something goes wrong, you search for a solution immediately.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Whenever there is a chance to get actively involved, you take it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You take initiative immediately even when others don't.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You use opportunities quickly in order to attain your goals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Usually you do more than you are asked to do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You are particularly good at realizing ideas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You use opportunities quickly in order to build up a business network.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other people rarely ask me about websites.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My opinions influence what types of websites other people visit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My opinion on websites seems not to count with other people.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Hint: how about Googling for “resource generator” and/or “position generator” “to find out more about the questions below?”

Position generator (order randomized)

[position1] through [position30]

(NB in the datafile you see which item is which number)

Please find below a list of professions. For every profession, please indicate whether you know someone who exercises this profession.

Note: by "knowing" we mean that you would greet this person and would call him or her by their first name.

		acquaintance	friend	family
teacher	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
mechanic	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
cook	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
unskilled labourer	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
lawyer	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
scientist	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Resource generator (order randomized)

[resource*] variables (NB only coded so that 1=yes,0=no)

Do you know anyone who...

Note: by "knowing" we mean that you would greet this person and would call him or her by their first name.

		acquaintance	friend	family
can do your shopping when you (and your household members) are ill	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
can give advice on matters of law (problems with landlord, boss, or municipality)	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
knows a lot about governmental regulations	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
can repair a car, bike, etc.	<input type="radio"/> no <input type="radio"/> yes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Who knows who?

[k1] through [k65] (order is randomized)

Please indicate to what extent you know the following students. It's a long list: stay focused!

	Don't know	Know who (s)he is	Acquaintance	Friend	Close friend
Venema J. (Joost)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Munoz Alcantara J. (Jesus)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tielen M.A.J. (Marco)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Heuvel T.T.P. van den (Tim)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bets A. (Arturo)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nelson N.M. (Nicolas)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The network of financial resources

Not as separate variables in the data set: contacts with others who can help you financially. In the data set you do find:

[numcontacts] = number of mentioned contacts

Suppose you had an interesting business idea, but for this you needed an amount of money that was beyond your own financial capabilities.

Which persons (if any) **in your family** could you ask to borrow such an amount from? Please just write down their first names and initial of second name.

1.
2.
3.
4.
5.

Which persons (if any) **in your network of friends and acquaintances** could you ask to borrow such an amount from? Please just write down their first names and initial of second name.

1. I've already mentioned this person before
2. I've already mentioned this person before
3. I've already mentioned this person before
4. I've already mentioned this person before
5. I've already mentioned this person before

Which persons (if any) **at the university** could you ask to borrow such an amount from? Please just write down their first names and initial of second name.

1. I've already mentioned this person before
2. I've already mentioned this person before
3. I've already mentioned this person before
4. I've already mentioned this person before
5. I've already mentioned this person before

Not as separate variables in the data: characteristics of contacts. In the data set you do find:

[propfemalecontacts] = proportion of female contacts

[avespeakcontacts] = average speaking terms with contacts

[aveagecontacts] = average age category of contacts (0/1/2/3)

Note: about half of you got this question in a different format, more graphical format, but this does not make a difference in terms of the responses

We'd like to find out a little more about the contacts you mentioned.

	b	a	c
What is the age of this contact?	<input type="radio"/> 15-25 <input type="radio"/> 26-40 <input type="radio"/> 41-60 <input type="radio"/> 60+	<input type="radio"/> 15-25 <input type="radio"/> 26-40 <input type="radio"/> 41-60 <input type="radio"/> 60+	<input type="radio"/> 15-25 <input type="radio"/> 26-40 <input type="radio"/> 41-60 <input type="radio"/> 60+
Gender of the contact?	<input type="radio"/> Male <input type="radio"/> Female	<input type="radio"/> Male <input type="radio"/> Female	<input type="radio"/> Male <input type="radio"/> Female
How often do you speak with this contact?	<input type="radio"/> Daily <input type="radio"/> Weekly <input type="radio"/> Monthly <input type="radio"/> Less often than monthly	<input type="radio"/> Daily <input type="radio"/> Weekly <input type="radio"/> Monthly <input type="radio"/> Less often than monthly	<input type="radio"/> Daily <input type="radio"/> Weekly <input type="radio"/> Monthly <input type="radio"/> Less often than monthly

Next

Not as separate variables in the data set: for all mentioned contacts, define the relation between all pairs of contacts. In the data set you do find:

[dontknowmatrix] = Number of don't know answers in matri

[sumofmatrix] = sum of tie strengths between contacts

[nummatrixentries] = number of matrix entries filled out

[density] = density of ties among the contacts (when known)

Part 1 – startup analyses

STARTUP ANALYSES ON THE NETWORK DATA

1.0) Create Ucinet data based on the SPSS data. Use the [k?] variables. Consult the Ucinet manual on how to read in a complete network matrix. Make sure to include the [id] variable, because this is the variable that will allow you to merge any network characteristics that you might calculate in Ucinet back to the SPSS data.

1.1) Have a look at the network data in Ucinet. The connections have values that run from 1 to 5, where (1=don't know, 2=know name, 3=acquaintance, 4=friend, 5=closefriend). There are zeroes on the main diagonal, and some empty values.

This is against standard Ucinet policy, where typically a zero represents “no connection”. Change the matrix in such a way that all values of 1 through 5 decrease by 1 (so that afterwards they vary from 0 to 4). You can change the main diagonal values as well, if you want. Also, consider what you want to do with the missing values. Leave them empty? Give them a value? Something else? Explain what you do and how you do it.

1.2) Create a meaningful drawing of the network in NetDraw so that we can get an idea about what the network looks like, and calculate the density of the network. Include at least one characteristic of the node (for instance, gender or age) in the picture, and use weighted lines between the dots. Elaborate on what we see in the picture.

1.3) The data are from three groups of students: those who follow the course as part of their Master Innovation Sciences, those who follow the course as part of their electives for an IE Master, and those who follow the course as part of the university wide “social entrepreneurship” program. Run a sensible clique analysis, and try to find these three groups back as good as you can (or: show that it cannot be done). Explain your strategy, and make sure to compare your answers to the variable [coursetype] in the data.

STARTUP ANALYSES ON THE SURVEY DATA

2.1) Have a look at the survey data. We are supposed to have the complete answers of 70 students. How many do we have? Check the data for missing values and other weird stuff. Where appropriate, repair mistakes. (If you find really weird stuff – send me an email.)

2.2) The sets of variables [aboutyou*], [position*], and [resource*] are similar in the sense that they try to measure an underlying characteristic or characteristics of the person through a larger set of items.

a. Consider the variables [aboutyou*]. Do these variables form a scale? Can the variables be sensibly reduced to fewer dimensions (use factor analysis!). Which dimensions might that be? Reduce the data to either a single or several dimensions, create these dimensions in your data, and label the dimensions appropriately.

b. The same for the [position*] variables.

c. The same for the [resource*] variables.

Note: to keep your SPSS data set a bit more tractable, you could now remove the separate items for the [aboutyou*], [position*], and [resource*] variables if you want, and just keep the dimensions you just created.

Part 2 – A paper connecting the network data to the survey data

(2/3 of your grade for this assignment)

The second part of the assignment is to write a paper, just like you did for the first assignment. The topic of your paper you can decide for yourself, *as long as it connects network characteristics (from your Ucinet file) to the survey data*. Use answers to the previous questions whenever this is useful. You do not have to use all your answers from the first part for the paper though.

It is probably easiest to *first come up with at least two sensible hypotheses that connect the survey data to the network data*. Typically, that would be a statement of the kind “students with a higher X will tend to have a lower/higher Y”, but other possibilities exist as well: you could look at differences between groups, boy/girl differences etc. Be original – don’t take the easiest way out – and explain in detail why these hypotheses make sense (how about using some literature that we covered in class?).

Structure your paper as follows (similar to Assignment 1). The number of pages mentioned is a rough guideline:

INTRO (max 1 page)

Briefly present the subject you chose. Don't start with "For the network course we had to write a paper ..." as the intro to your work. Think about why what you do might be of scientific interest. End the introduction with your main research question. Make sure to include proper references to papers that you know from the course (or somewhere else).

THEORY AND HYPOTHESES (2-4 pages)

Present your theory and hypotheses here. Do not to be too casual in this part; make use of the theories that were presented earlier in the course to deduce hypotheses, and to decide which of the network measures you choose to calculate. Carefully explain why you think that your theory makes sense. Where appropriate, refer in detail to the right scientific papers (e.g., by Burt, Coleman, or others).

DATA AND RESULTS (3-5 pages)

Explain how you are going to test the hypotheses: which network characteristics are you going to calculate/use and which kind of analysis are you going to use to test the hypotheses? Explain how you calculate the characteristics. If necessary, add syntax to your answer. There is no need to introduce the data and data collection. Just mention that we use the network course data.

Test the hypotheses. Use an alpha of 10%. Explain and interpret your finding. Note that given the relatively low sample size, chances are that you will not find statistically significant results, unless the effects you are trying to find are really large.

CONCLUSION AND DISCUSSION (2 pages)

Present the conclusion. What do the results say about your theory? Would it make sense to run other analyses (explain which ones and why)? How could one improve upon this analysis in future analyses? How valid do you think the results are? How reliable are the results? Can the results be generalized beyond this specific research? Explain your answers. Finally, explain what the reader can/should learn from your paper.

Be complete and make this as close to a real paper as possible, including referencing, proper tables and figures, etc.

Bonus points are for those who make more of it. For instance, you could analyze separately on subsets of the sample, run more complicated analyses, use variables in different ways (transformations anyone?), spend some quality time on the reliability of the data (some of you made a bit of a mess while inputting) or do something else that impresses me. Note that analyses at the level of multiple regression and the like are assumed to be part of your standard repertoire.

If you have any questions, let me (CS) know at c.c.p.snijders@gmail.com. If I do not reply within 24 hours, I have probably missed or forgot about your mail – please resend it. Do not hesitate to ask my help, but keep your questions within reason (so don't ask me how to get your data into Ucinet... this you have to figure out yourself).

Logistics

Answers can be written in groups of **at most two** but if you prefer to do it by yourself this is allowed as well. Obviously, papers handed in should be original, not a copy of someone else's. Refer to others whenever appropriate. *There is no mercy for those plagiarizing.*

Deadline: December 5, 23:59.

Deliverable: Deliver as a *pdf*-attachment in the email box of Uwe Matzat (umatzat@gmail.com) **and** Chris Snijders (c.c.p.snijders@gmail.com). If you do not know how to create a pdf, ask your fellow students. NB Give your file a unique and insightful name, for instance <name1><name2>_0zm05_asgn2.doc. In any case, don't name it "assignment.doc".

If – for some *good* reason – you are unable to meet the deadline, make sure to mention this by email to me (CS) long before the deadline expires.

Best regards, good luck,

Chris Snijders

HINTS, TIPS, AND ISSUES TO CONSIDER

On the first part

On recoding: don't just delete everything that looks suspicious! Data are precious and should only be deleted if you are certain they must go. Think about trying to reconstruct data for those cases where you have missing (or weird) values.

On predicting the course types: Don't just say "you can see 5 cliques by looking at the graph". Define them properly – the reader should not need to make conclusions based on eyeballing the data. Be specific on the kind of method you use, and do come up with a conclusion at the end: is it possible to predict the course types well or not?

The aboutyou / position / resource data: use the hints as supplied in the assignment (that is: read and use the suggested literature)! Do remember that the data are real, and matters might have turned out badly for some reason. Perhaps not everything that you hoped to be able to do, is possible. On the other hand, don't just say "does not work well" without giving it a proper try.

On the second part

Don't make your main question too broad ("How can social network characteristics be connected to personal characteristics?").

If you symmetrize your data: explain how and why! Also, check your data on how often actual symmetry occurs.

Do not use data to derive your hypotheses. You first derive hypotheses and then test them with the data.

Refer to the literature, but wisely. Don't just refer to Granovetter because at some point you have accidentally mentioned the phrase "weak ties". Make your arguments explicit.

The theory part should consist of theoretical arguments *as applied to our classroom data*: don't just copy slide content.

It will probably be hard to find significant effects in your analyses: the group is too small. Still: it requires some effort to do the data justice.

Check the assumptions of your analyses and show it (don't just mention that you did it).

How about outliers? Especially in small data sets, outliers are a real threat.

Don't just show SPSS output. Have a look at what a regression table looks like in a real paper, and copy that layout.