Individual assignment

Description of the data sets

1 Task

The individual assignment serves to assess the achievement of the overarching goal of the course: "By the end of the course, you should be able to develop hypotheses that relate to the structure and dynamics of (social) networks, and apply statistical network models to test those hypotheses".
The assignment consists of a data analysis presented in a report.

1. We provide you with a few data sets described in Section 2. The data sets are simulated data derived from real case studies. Read the description and choose the data set that you are going to analyse.
   Please, enter your choice using the item "Choice of the data set" in Moodle by Tuesday, December 14th

2. Given the chosen data set, develop hypotheses that relate to the structure/dynamics of the chosen network(s). You can look for one or two research papers that work on a similar problem to get inspired. If you do so, please include the references to those papers in the report.

3. Perform the analysis to test those hypotheses: choose and apply the adequate network model

4. Write a report (2,500 words max.) illustrating your analysis. You find a template for the report in the file ReportStructure.pdf. Feel free to modify the structure at your convenience. When you write the report, imagine that you are addressing it to an audience that might not be familiar with statistical network models (i.e., argue for and briefly describe the adequate network model, illustrate the different steps of the analysis, comment on the results and difficulties encountered, etc.)

5. Submit the report using Moodle. The report should be submitted along with the R script as an R or Rmd file. If you use an Rmd file to write the report, please attach the corresponding pdf file. The deadline for submission is January, 14th at noon

On Monday, December 20th, we will not have a lecture, but we will have an online office during which you can ask questions about the data set and your analysis. You can book your time slot using the option in Moodle.

2 Description of the data sets

2.1 Advice

The folder `advice.zip` contains data about advice ties among bachelor university students in a cohort. The data were collected over three years, at the end of each academic year, through the question "Consider the last academic year. Could you name the other students in the cohort whom you recurrently consulted for help and support on course-related tasks?". The relational data are represented by the three adjacency matrices `net1`, `net2` and `net3` collected at the first, second and third time point, respectively.
The file `attributes.csv` contains the information on the following students' characteristics:

- id: student identifier
- gender: 1 = male, 2= female
- age: in years
- workexp: 1 = non student-worker, 2 = student-worker
- nationality: 1 = Italian, 2 = not-Italian

2.2 AdviceGrade

The folder `advice.zip` contains data about advice ties among bachelor university students in a cohort. The data were collected over three years, at the end of each academic year, through the question "Consider the last academic year. Could you name the other students in the cohort whom you recurrently consulted for help and support on course-related tasks?". The relational data are represented by the three adjacency matrices `net1`, `net2` and `net3` collected at the first, second and third time point, respectively.
The file `attributes.csv` contains the information on the following students' characteristics:

- id: student identifier
- gender: 1 = male, 2= female
- age: in years
- workexp: 1 = not student-worker, 2 = student-worker
- nationality: 1 = Italian, 2 = not-Italian

The file **grade.csv** contains the average grade obtained by the students over each academic year. The university grades in Italy vary between 0 and 30, with 18 being the grade necessary to pass an exam. The average grade was collected as a measure of student performance.

2.3 Coauthorship.zip

The folder `coautorship.zip` contains the information on co-authorship ties between researchers in a small community of physicists who published their first paper in 2005. The files in the folder are:

- `coautorship.csv`: a valued network. The value in the cell x_{ij} is the number of papers co-authored by physicists i and j and published from 2010 to 2018
- `coautorshipPrev.csv`: a valued network. The value in the cell x_{ij} is the number of papers co-authored by physicists i and j and published from 2005 to 2009
- The file `attributes.csv` contains the information on the following physicists' characteristics:
  - Id: physicist identifier
  - physics: the research field (1=theoretical physics, 2=applied physics)
  - gender: 1=male, 2=female

2.4 Collaboration.zip

The folder `collaboration.zip` contains information on collaboration tie among 75 companies. Two companies are connected if they collaborated on at least one project within the six months before the data collection. The file `attributes.csv` contains the information on the following companies' characteristics:

- id: company identifier
- size: number of employees
- location: location of the headquarter of a company (1=US, 2=Europe, 3 =elsewhere)
- sector: the sector in which a company operates (1= extraction of natural resources, 2=agriculture)

2.5 Communication.zip

The file `communication.csv` contains the information on communication ties between the employees of a company. Data were collected using a survey and the question "With whom did you have to communicate to complete your work effectively over the past three months?". The file `attributes.csv` contains the information on the following employee characteristics:

- Id: employee identifier
- Floor: the floor on which an employee's office is located (1= first floor, 2= second floor, 3 = third floor)

- Seniority: binary variables indicating whether an employee has worked more or less than 10 years in the field in which the company is operating (1=senior ($\geq$ 10), 0 = junior (<10)
- Projects: number of projects in which an employee has been involved with

2.6 Exchange.zip

The folder `exchange.zip` contains information on money transfers between 65 banks. The files in the folder are:

- exchange06.csv: a valued network. The value in the cell x_{ij} is the amount of money (in billions) transferred from bank i to bank j in 2005
- exchange07.csv: a valued network. The value in the cell x_{ij} is the amount of money (in billions) transferred from bank i to bank j in 2007
- exchange08.csv: a valued network. The value in the cell x_{ij} is the amount of money (in billions) transferred from bank i to bank j in 2008
- The file `attributes.csv` contains the information on the following bank characteristics:
  - Id: bank identifier
  - continent: the continent in which the head quarter of the bank is (1=Asia, 2=Europe, 3=America, 4=Africa, 5=Australia)
  - size: number of employees in thousands

2.7 HospitalPerformance

The folder `HospitalPerformance.zip` contains information on patient transfer among 74 hospitals of a region collected at three time points. The relational data are represented by the three adjacency matrices `net1`, `net2` and `net3` collected at the first, second and third time point, respectively. Two hospitals are connected by a patient referral event (x_{ij}) when a sender hospital refers a patient to a receiving hospital. Because patient referral typically requires a considerable level of coordination and communication between partner hospitals, patient referral events are indicators of the presence of an underlying collaborative relation between the hospitals involved. The file `attributes.csv` contains the information on the following hospitals' characteristics:

- id: hospital identifier
- private: binary variables indicating whether an hospital is public (0) or private (1)
- unit: membership of hospitals to the different administrative units in which the region is partitioned

– nbeds1, nbeds2, nbeds3: size of the hospital measured by the number of beds. The number of beds can vary over time and was collected at the three observational time points

The file `geodist.csv` contains the information on the logarithm of the driving distance (measured in Km) between two hospitals.
The file `performance.csv` contains the information on a performance measure of the hospitals. The performance is measured on an ordinal scale, with 1 indicating the lowest performance, and 4 the highest performance

2.8 Patient

The folder `patient.zip` contains information on patient transfer among 74 hospitals of a region collected at three time points. The relational data are represented by the three adjacency matrices `net1`, `net2` and `net3` collected at the first, second and third time point, respectively. Two hospitals are connected by a patient referral event (x_{ij}) when a sender hospital refers a patient to a receiving hospital. Because patient referral typically requires a considerable level of coordination and communication between partner hospitals, patient referral events are indicators of the presence of an underlying collaborative relation between the hospitals involved. The file `attributes.csv` contains the information on the following hospitals' characteristics:

– id: hospital identifier

– private: binary variables indicating whether an hospital is public (0) or private (1)

– unit: membership of hospitals to the different administrative units in which the region is partitioned

– nbeds: size of the hospital measured by the number of beds. The number of beds can vary over time and therefore was collected at all three time points

The file `geodist.csv` contains the information on the logarithm of the driving distance (measured in Km) between two hospitals.

2.9 Support.zip

The file `support.txt` contains the information on support ties among 11 years-old pupils in a classroom. The data were collected using the questions "Who did support you in difficult moments (e.g., when you fought with a friend, got a bad grade, mocked by others)?"
The file `attributes.csv` contains the information on the following pupils' characteristics:

– Id: id of the pupils

– gender: 0 = male, 1= female

– ethnicity: 1 = native, 1 = born in the country by migrant parents or arrived in the country before being 3 years old, 3 = arrived in the country after being 3 years old

2.10 Transfer.zip

The file `transfer.csv` contains information on patient transfer among the hospitals of a region. Two hospitals are connected by a patient referral event ($x_{ij} = 1$) when a sender hospital refers a patient to a receiving hospital. Because patient referral typically requires a considerable level of coordination and communication between partner hospitals, patient referral events are indicators of the presence of an underlying collaborative relation between the hospitals involved. The file `attributes.csv` contains the information on the following hospitals' characteristics:

- id: hospital identifier
- private: binary variables indicating whether the hospital is public (0) or private (1)
- unit: membership of hospitals to the different administrative units in which the region is partitioned
- nbeds: size of the hospital measured by the number of beds

The file `geodist.csv` contains the information on the logarithm of the driving distance (measured in Kilometers) between two hospitals