TARAS SHEVCHENKO
NATIONAL UNIVERSITY OF KYIV
Faculty of Sociology
Department of Methodology and Methods of Sociological Research

Yuriy Savelyev

# Social Networks Analysis

## Learning package

for students

| | |
|---|---|
| branch | 05 Social and behavioral sciences |
| specialty | 054 Sociology |
| education level | master degree |
| education program | «Social technologies» |
| | «Gender studies» |

| | |
|---|---|
| *ECTS* | *5* |
| *Language of teaching learning and evaluation* | *English* |
| *Form of final control* | *exam* |

KYIV – 2020

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет соціології
Кафедра методології та методів соціологічних досліджень

Ю.Б. Савельєв

# Аналіз соціальних мереж

Навчально-методичний комплекс

для студентів

| | |
|---|---|
| галузь знань | 05 Соціальні та поведінкові науки |
| спеціальність | 054 Соціологія |
| освітній рівень | магістр |
| освітня програма | «Соціальні технології»<br>«Гендерні студії» |

| | |
|---|---|
| *Кількість кредитів ECTS* | *5* |
| *Мова викладання, навчання та оцінювання* | *англійська* |
| *Форма заключного контролю* | *екзамен* |

КИЇВ – 2020

УДК 316.3

Укладач:
Савельєв Ю.Б., д. соц. н., доц.

Навчально-методичний комплекс з дисципліни «Аналіз соціальних мереж» розроблений із використанням кредитно-модульної системи організації навчального процесу і призначений для студентів факультету соціології спеціальності 054 «соціологія», освітній рівень «магістр».

Рецензенти:

Дембіцький С.С., доктор соціологічних наук, провідний науковий співробітник відділу методології і методів соціології Інституту соціології НАН України

Сальнікова С.А., кандидат соціологічних наук, доцент кафедри методології та методів соціологічних досліджень факультету соціології Київського національного університету імені Тараса Шевченка

Схвалено на засіданні кафедри методології та методів соціологічних досліджень факультету соціології (протокол № 12 від 10 березня 2020 р.)

Схвалено навчально-методичною комісією факультету соціології (протокол № 6 від 22 квітня 2020 р.).

Схвалено і рекомендовано до видання Вченою радою факультету соціології Київського національного університету імені Тараса Шевченка (протокол № 9 від 24 квітня 2020 р.)

Савельєв Ю.Б.
Аналіз соціальних мереж: Навчально-методичний комплекс з дисципліни для студентів спеціальності 054 «соціологія», освітній рівень «магістр». - К.: Київський національний університет імені Тараса Шевченка, 2020. – 52 с.

Savelyev Y. Social network analysis: Learning package for students in specialty 054 Sociology, master education level. - K.: Taras Shevchenko National University of Kyiv, 2020. – 52 p.

CONTENTS

1. Course description…….. 6
Learning outcomes…….....7
Evaluation and grading…… 9
2. Course structure and thematic plan……. 10
3. Course assignments for seminars…… 11
Module 1. Analysis of networks in research of social processes……..11
Topic 1. Theoretical concepts and methodological
foundations of SNA…… 11
Questions for discussion …..……..... 11
Practical tasks……..11
Recommended literature…… … 11
Topic 2. Network Characteristics & Metrics …….....12
Questions for discussion…… 12
Practical tasks…… 12
Recommended literature…….. 12
Topic 3. Measures of a Position in the Network…….. 13
Questions for discussion…… 13
Practical tasks…… 13
Recommended literature……..... 13
Module 2. Collecting, processing and analysing network data…….14
Topic 4. Research Design and Data Collection
in Network Analysis……14
Questions for discussion…… 14
Practical tasks…… 14
Recommended literature…….. 15
Topic 5. Software tools & Data management
in Network Analysis……16
Questions for discussion…… 16
Practical tasks……..16
Recommended literature…….. 16
Topic 6. Methods of Analysis of Network Data…….. 17
Questions for discussion…… 17
Practical tasks…… 17
Recommended literature…….. 18

4. Testing and evaluation…… 19
Module 1. sample test…….. 19
Module 2. sample assignment…… 22
5. Requirements for analytical report of individual network research project…….. 24
6. Exam questions sample……. 25
7. Course lectures outline…….. 27
Module 1. Analysis of networks in research of social processes……. 27
Module 2. Collecting, processing and analysing network data…… 38
8. Course literature and resources…….. 50

## 1. Course description

The 5 ECTS course "Social network analysis" is designed for full-time training sociology students. The language of instruction is English and the course is scheduled for the first semester of master degree program in sociology.

**The course goal** – students' comprehension of basic principles and ability to use social network analysis (SNA) to advance their research competencies.

**Prerequisites to student's knowledge and skills:**

*1. Knowledge of basic methods of data collection and analysis*

*2. Computer intermediate skills*

*3. English level to read professional literature and comprehend lecture materials*

**The course summary**:

Methods of analysis of social networks are necessary to identify and understand the structural relationships between different actors (individuals, organizations, countries, etc.) in contemporary society. These methods are of particular importance along with development of Internet communications and the global spread of social networks. The course aims to introduce research potential, theoretical and methodological foundations of social network analysis methods and develop basic skills to design and conduct network analysis in practice using programing environments Pajek, Gephi, R or UCINET.

**The course learning objectives**:

- Knowledge of essential SNA concepts, measures, methods and SNA potential in research of social interactions
- Knowledge and skills to design research of networks and choose appropriate SNA methods
- Basic skills of collecting and processing network data
- Basic skills of analysis of networks in programming environments Pajek, Gephi, R or UCINET

This aims at developing students' competencies:

• Information and communication technology skills (zk12)
• Ability to justify the use of the latest methods of collecting and analyzing sociological information for solving practical problems in different spheres of public life (fk11)
• Ability to diagnose the state of communication and organizational processes, to develop and justify directions of increasing their efficiency (fk 14)
• Ability to independently use knowledge of modern sociological theory and methodology to solve problems of applied research of social communities, institutes, processes and public opinion (fk15)

**Learning outcomes (LO):**

| LO code | Learning outcome | Forms or methods of teaching | Methods and criteria of evaluation | Proportion of finale grade |
|---|---|---|---|---|
| *1.1* | Knowledge of essential SNA concepts, measures, methods and SNA potential in research of social interactions | Lecture, seminar, individual work | *test with answer options and open answers* | *20* |
| *1.2* | Specificity of social network research planning and peculiarities of using network analysis methods | Lecture, seminar, individual work | *test with answer options and open answers; individual social network research project* | *10* |
| *2.1* | Plan, select appropriate data collection and analysis methods, and conduct social network research | Lecture, seminar, individual work | *test with answer options and open answers; individual social network research project* | *20* |
| *2.2* | Perform of network data collection and analysis | Lecture, seminar, individual work | *practical assignment test; individual social network research project* | *20* |
| *2.3* | Perform analysis of networks in programming environments Pajek, Gephi, R or UCINET | Lecture, seminar, individual work | *practical assignment test; individual social network research project* | *30* |

**Relation of the learning outcomes (LO) of the discipline to the overall education program outcomes:**

| Learning outcomes / Program outcomes | 1.1 | 1.2 | 2.1 | 2.2 | 2.3 |
|---|---|---|---|---|---|
| Apply the latest sociological theories and methods of analysis of social phenomena and processes (prn8) | + | + | + | + | + |
| To use modern information and communication technologies in professional communication; exchange of information; collecting, analyzing, processing and interpreting data and visually presenting the results of scientific activities (prn11) | + | + | + | + | + |
| To use the latest methods of collecting and analyzing sociological information for solving practical problems (prn17) | + | + | + | + | + |
| Demonstrate the ability to critically analyze theoretical, methodological and practical professional literature (prn20) | + | + | | | |
| To carry out sociological analysis of current socio-political, economic and cultural events, processes and phenomena of the present (prn19) | + | + | | | + |
| Conduct comparative studies and draw conclusions about the nature and causes of similarities and differences of different societies (prn21) | + | + | + | + | + |

**Evaluation and grading.**

**- Semester evaluation:**

1. Practical tasks for topic 1, topic 2, topic 3, LO 1.1, LO 1.2 = 10 points *(maximum)* / 2 points for each task
2. Module control test 1 (MCT 1) from topic 1, topic 2, topic 3, LO 1.1, LO 1.2 = 10 points *(maximum)* / 6 points *(minimum to pass)*
3. Practical tasks for topic 4, topic 5, topic 6 LO 2.1, LO 2.2, LO 2.3 = 10 points *(maximum)* / 2 points for each
4. Module control test 2 (MCT 2) of theme 4, theme 5, theme 6, LO 2.1, LO 2.2, LO 2.3 = 10 points *(maximum)* / 6 points *(minimum to pass)*
5. Final analytical report on social networks research and presentation of results, LO 1.1, LO 1.2, LO 2.1, LO 2.2, LO 2.3 = 20 points *(maximum)* / 12 points *(minimum to pass)*

**- Final evaluation***:*

*Exam = 40 points (maximum) / 24 points (minimum to pass)*

**Conditions for admission to the final evaluation:** The student is admitted to the exam if successful completion and personal presentation (not below the threshold level of positive assessment) of practical tasks (topics 1, 2, 3, 4, 5, 6 = 12 points and above), two module control tests (MCT 1 = 6 points and above; MCT 2 = 6 points and above), which must be submitted in writing no later than the deadline provided for in the thematic plan, as well as the successful completion of the individual research project (12 points and above) by personally submitting the final individual social network research report in writing and electronically with mandatory network data file and presentation of the report within the deadline provided thematic plan.

**Organization and timeline of evaluation**:

1. Module control test 1 (MCT 1) on topic 1, theme 2, theme 3, LO 1.1, LO 1.2, written in a practical session - after topic 3
2. Module control test 2 (MCT 2) on topic 4, topic 5, topic 6, LO 2.1, LO 2.2, LO 2.3, written in a practical session - after topic 6
3. The final analytical report on research of social networks, LO 1.1, LO 1.2, LO 2.1, LO 2.2, LO 2.3 shall be submitted personally in writing and in electronic form with the required network data file before the deadline, the presentation of the reports shall be held at practical activity - after topic 6.

**Grading:**

| | |
|---|---|
| Excellent | 90-100 |
| Good | 75-89 |
| Satisfactory | 60-74 |
| Fail | 0-59 |

## 2. The course structure and thematic plan

| № | Topic | Work hours | | |
|---|---|---|---|---|
| | | lectures | seminars | students' work |
| **1. Analysis of networks in research of social processes** | | | | |
| 1 | 1. Theoretical concepts and methodological foundations of SNA | 4 | 2 | 10 |
| 2 | 2. Basic properties and metrics of social networks | 6 | 4 | 10 |
| 3 | 3. Centrality measures in networks | 4 | 4 | 10 |
| 4 | *Module control test 1* | | | 10 |
| **2. Collecting, processing and analysing network data** | | | | |
| 5 | 4. Research design, network data collection and processing | 4 | 6 | 10 |
| 6 | 5. Analysis of networks in programing environments Pajek, Gephi, R or UCINET | 4 | 4 | 10 |
| 7 | 6. Methods of analysis of network data | 2 | 6 | 10 |
| 8 | *Module control test 2* | | | 10 |
| 9 | *Final analytical report of individual network research project and presentation* | | | 20 |
| | **Overall** | **24** | **26** | **100** |

**Total workload** ***150*** *hours,* including:
Lectures **-** ***24*** *hours*
Seminars **- 26** *hours*
Individual work **-** ***100*** *hours*

## 3. Course assignments for seminars

### Module 1. Analysis of networks in research of social processes

### Topic 1. Theoretical concepts and methodological foundations of SNA

Before the class students must read the recommended literature and prepare practical tasks.

**Questions for discussion**:
1. Historical and Theoretical Foundations of SNA
2. SNA Application in Social Research. Which sociological theories relate to SNA?
3. Stanley Milgram's experiments and Grannovetter's idea on weak ties
4. What SNA tells about the Rise of the Medici?
5. Peculiarity of Network data. Main types of variables
6. Identify and describe different levels of analysis in SNA. What is SNA capable of?

**Practical Tasks**:
1. Find 2 SNA research examples (with references). Describe theory, methodology and the findings of your examples.
2. Formulate own example of a problem and a research question suitable for SNA.
3. Suggest 3 examples of variables for SNA. Provide arguments of your choice.

**Questions for self-control:**
*1. What is an advantage of network approach in social research?*
*2. Which sociological theories relate to SNA?*
*3. What types of variables exist in network data?*

*Recommended literature:*

- Wasserman S., Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994. P. 2-66.
- Borgatti S., Everett M., Johnson J. Analyzing Social Networks. London: SAGE, 2013. 1. Introduction.
- Scott J. Social network analysis. 4th edition. London: Sage, 2017. 1. What is Social Network Analysis? 2. The History of Social Network Analysis.
- Freeman Linton (2004). The development of social network analysis. *A Study in the Sociology of Science*.
- Granovetter 1973 The Strength of Weak Ties. AJS Volume 78 Number 6 (*Грановеттер* Сила слабых связей // Экономическая социология Т. 10. № 4. 2009)
- Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American journal of sociology*, *98*(6), 1259-1319.

## Topic 2. Basic properties and metrics of social networks

Before the class students must read the recommended literature and prepare practical tasks.

**Questions for discussion**:

1. Network terminology. How can we represent networks?
2. Define and compare different types of networks
3. Define and compare Network Distance Measures: Diameter, Path, Walk, Cycle, Geodesic distance, Average Path length
4. Network Structure. How can we estimate network structure?
5. Network Cohesion: Density, Transitivity. Clustering. Network Density vs. Clustering coefficients.
6. Degree Distributions in Networks. Why Degree Distributions are important?
7. Random Graphs Models and their application in Network Analysis
8. How can we use Theorem on Network Structure?

**Practical Tasks**:

Make groups of 2 students each and:

1. Calculate possible number of networks for 5 nodes.
2. Draw 1 undirected, 1 directed and 1 weighted network with minimum 5 nodes each and write representation of each network as matrices, edge lists and adjacency lists. Compare different representations.
3. Identify paths, walks, geodesics, components and diameter in these networks.
4. Calculate mean nodal degree and explore degree distribution for these networks.
5. Calculate density, global and local clustering coefficients for these networks.

**Questions for self-control:**

*1. What are the main types of networks?*
*2. What are the main network characteristics?*
*3. What are the key differences between random and observed networks?*

*Recommended literature:*


- Wasserman S., Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994. P.67-165, 243-248.
- Borgatti S., Everett M., Johnson J. Analyzing Social Networks. London: SAGE, 2013. 2. Mathematical foundations. 9. Characterizing whole networks.
- Hanneman R., Riddle M. Introduction to Social Network Methods. 2005. 3. Using graphs to represent social relations. 5. Using matrices to represent social relations. 7. Connection and distance.

## Topic 3. Centrality measures in networks

Before the class students must read the recommended literature and prepare practical tasks.

**Questions for discussion**:

1. Explain the concept of Centrality. Why and how do we use this concept in SNA?
2. Compare Centrality Measures: Degree Centrality, Betweenness Centrality, Closeness Centrality, Decay Centrality, Eigenvector Centrality, Bonacich Centrality.
3. Which Centrality Measures do we use for which purposes?
4. How do we interpret Centrality Measures?
5. Explain the limitations and constraints of various centrality measures in undirected, directed and weighted networks.

**Practical Tasks**:

Make groups of 2 students each and:

1. Suggest 2 real world examples of networks: 1 undirected &1 directed with minimum 5 nodes each.
2. Compute Degree Centrality, Betweenness Centrality, Closeness Centrality, Decay Centrality for at least 2 nodes from each network.
3. Interpret the obtained centrality measures and describe a position of selected actors in the network.

**Questions for self-control:**

*1. How can we measure position of a node in the network?*

*2. What does centrality mean?*

*Recommended literature:*


- Wasserman S., Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994. P.169-215.
- Borgatti S., Everett M., Johnson J. Analyzing Social Networks. London: SAGE, 2013. 10. Centrality.
- Hanneman R., Riddle M. Introduction to Social Network Methods. 2005. 10. Centrality and power.


**End of Module 1**

***Module control test 1 - MCT 1 (written test)* = max 10 / min 6**

## Module 2. Collecting, processing and analysing network data

### Topic 4. Research design, network data collection and processing

Before the class students must read the recommended literature and prepare practical tasks.

**Questions for discussion**:

1. Compare Research design in SNA and other types of sociological research.
2. Types of SNA Research design. Explain capacity and limitations of these types.
3. Features of design to study ego networks.
4. Explain problems of sampling in SNA.
5. What are the main strategies of sampling in SNA?
6. Identify types, sources and boundaries of network data.
7. Measurement in SNA.
8. Compare data collection in SNA and other types of sociological research.
9. Data collection methods in SNA. Compare their advantages and limitations.
10. Explain problems of creating questionnaires for network data collection.

**Practical Tasks**:

1. Identify and describe research design for 2 SNA research examples (from Class 1, task 1).
2. Suggest research design, a plan for sampling and network data collection for your own SNA research project. Provide arguments of your choice.
3. Suggest examples of structural and composition variables for your own SNA research project. Explain what you are going to measure and why.
4. Find 1 example of research (with reference) on cognitive social networks (cognitive social structure).
5. Find 1 example of research (with reference) on ego networks.
6. Prepare a draft of questionnaire including name generator, position generator, resource generator and name interpreter.
7. Collect data to describe your own ego networks of on-line and off-line social communication (e.g. during the recent week). Draw these networks by hand.

**Questions for self-control:**

*1. What are specific designs in research of social networks?*
*2. Why random sampling does not work well in research of social networks?*
*3. What are advantages of primary network data collection?*
*4. What are advantages of secondary network data?*
*5. What is the difference between name generator and position generator?*
*6. What is the difference between name generator and name interpreter?*
*7. By which criterion (or criteria) name generators differ from each other?*
*8. What are 4 strategies of sampling which can be used according to D.Knoke & S.Yang.*

*Recommended literature:*

- Wasserman S., Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994. P.28-58.
- Borgatti S., Everett M., Johnson J. Analyzing Social Networks. London: SAGE, 2013. 3. Research Design. 4. Data collection. 15. Ego networks.
- Knoke D., Yang, S. Social network analysis (Vol. 154). London: SAGE, 2008. 2. Network Fundamentals. 3. Data Collection.
- Scott J. Social network analysis. 4th edition. London: Sage, 2017. 3. Data Collection for Social Network Analysis.
- Models and Methods in Social Network Analysis. Cambridge: Cambridge University Press, 2005. 2. Recent Developments in Network Measurement. 3. Network Sampling and Model Fitting.

## Topic 5. Analysis of networks in programing environments Pajek, Gephi, R or UCINET

Before the class students must read the recommended literature and prepare practical tasks.

**Questions for discussion**:

1. Explain main features of data management in SNA.
2. Compare various software tools for SNA. Explain your choice
3. How to choose network data file format?
4. How to create network data file?
5. How to export and import network data files?
6. How to prepare network data for analysis?

**Practical Tasks**:

1. Install free software R, RStudio, Pajek (version 2.05) and Gephi on your computer.
2. Open data file "imports.net" (to be provided) with Pajek and Gephi.
3. Explore the data. Draw the network. What network is this?
4. Create your own network data file. What is the file format? Explain your choice.
5. Open your own network data file with Pajek and Gephi. Explore your data.
6. Open R via RStudio. Install igraph and statnet packages.
7. Open your data file with R (RStudio). Explore the data.

**Questions for self-control:**

*1. Which software tools provide the most advantages for network analysis?*
*2. What is the best way to create a network data file?*
*3. What are the key steps in network data management?*

*Recommended literature:*


- Borgatti S., Everett M., Johnson J. Analyzing Social Networks. London: SAGE, 2013. (5. Data Management.)
- Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj. Exploratory social network analysis with Pajek. 2nd ed. Cambridge: Cambridge University Press, 2011. P. 3-50, 369-382.
- Luke D. A Users Guide to Network Analysis in R. Springer, 2015. P.17-41.
- Hanneman R., Riddle M. Introduction to Social Network Methods. 2005. (5. Using matrices to represent social relations. 6. Working with network data.)
- UCINET 6 for Windows USER'S GUIDE. 2002. P.5-37. http://www.analytictech.com/ucinet/help.htm
- Gephi Tutorials: Learn how to use Gephi. URL: https://gephi.org/users/

## Topic 6. Methods of analysis of network data

Before the class students must read the recommended literature and prepare practical tasks.

**Questions for discussion**:

1. How to visualize networks correctly?
2. Analysis of Network properties.
3. Analysis of Cohesive Subgroups.
4. Analysis of Affiliations Networks.
5. Analysis of Equivalence.
6. Analysis of Ego Network data.

**Practical Tasks**:

1. Perform analysis of data file "imports.net" (from Class 5, task 2) with Pajek and Gephi: calculate average degree of the network, calculate diameter of the network, clustering coefficients, calculate the number of components of the network, calculate centrality measures. Interpret the results.
2. Open data file "florentine.dat" (to be provided) with R (via RStudio) and perform analysis of the network. Interpret the results.
3. Open a data file of your own SNA research project using Pajek and Gephi and examine the diameter and average clustering coefficient.
4. Using Pajek and Gephi generate random network with 25 nodes and examine the diameter and average clustering coefficient. Compare random network with your observed network. Interpret the results.
5. Open a data file of your own SNA research project with R (via RStudio) and perform analysis of the network. Interpret the results.

**Questions for self-control:**

*1. What are 4 categories of relations in the networks which can be studied according to Borgatti S., Everett M., Johnson J.?*
*2. How to estimate network properties?*
*3. How to analyze affiliation networks?*
*4. How to analyze ego networks?*
*5. What is the best way to visualize networks?*
*6. How to analyze cliques and subgroups in networks?*
*7. How to analyze structural equivalence?*
*8. What are 4 types of constraints which should be built into name generators according to D.Knoke & S.Yang?*

*Recommended literature:*

- Borgatti S., Everett M., Johnson J. Analyzing Social Networks. London: SAGE, 2013. (7. Visualization. 11. Subgroups. 12. Equivalence 13. Analyzing two-mode data. 15. Ego networks.)
- Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj. Exploratory social network analysis with Pajek. 2nd ed. Cambridge: Cambridge University Press, 2011. P. 57-133.
- Luke D. A Users Guide to Network Analysis in R. Springer, 2015. P.45-144.
- Hanneman R., Riddle M. Introduction to Social Network Methods. 2005. (8. Embedding. 9. Ego networks. 11. Cliques and sub-groups. 13. Measures of similarity and structural equivalence. 17. Two-mode networks.)
- Knoke D., Yang, S. Social network analysis (Vol. 154). London: SAGE, 2008. 3. Data Collection. 4. Basic Methods for Analyzing Networks. 5. Advanced Methods for Analyzing Networks.
- Wasserman S., Faust K. Social Network Analysis. Cambridge: Cambridge University Press, 1994. P. 250-343.

**End of Module 2**

***Module control test 2 - MCT 2 (practical assignments to demonstrate skills in network data analysis)* = max 10 / min 6**

**Presentation of individual network research project analytical report (during last class) = max 20 / min 12**

## 4. Testing and evaluation

### Module 1. sample test (tentative questions)

**1. Representation of a network in a matrix where the rows are actors-persons, and the columns are events which they attended is called:**

a) One-mode network
b) Directed network
c) Ego-centered network
d) Weighted network
e) Two-mode network
f) Partial network

**2. Is density higher in observed large networks or random graphs?**

_____
_____

**3. What type of network is this?**

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| A | 1 | 1 | 1 | 1 | 1 | 1 |
| B | 1 | 1 | 1 | 1 | 1 | 1 |
| C | 1 | 1 | 1 | 1 | 1 | 1 |
| D | 1 | 1 | 1 | 1 | 1 | 1 |
| E | 1 | 1 | 1 | 1 | 1 | 1 |
| F | 1 | 1 | 1 | 1 | 1 | 1 |

_____
_____
_____
_____

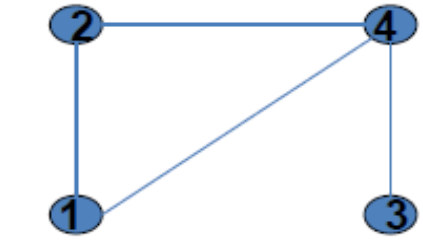
**4. What levels of Social Network Analysis would you identify?**

_____
_____
_____
_____

**5. Are random graphs more clustered than and observed large networks?**

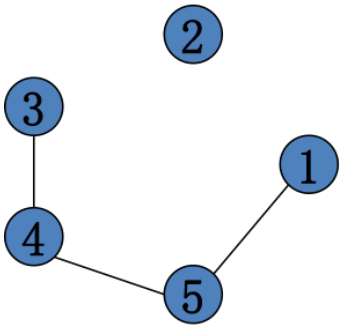
_____
_____

**6. What is global clustering coefficient for this network?**




_____

**7. What is local clustering coefficient for node 1?**

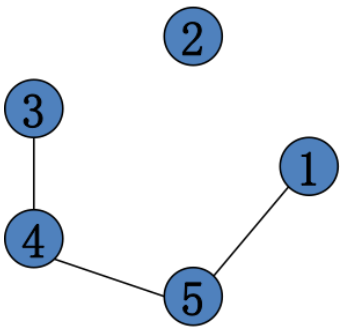



_____

**8. Any network can be represented as:**

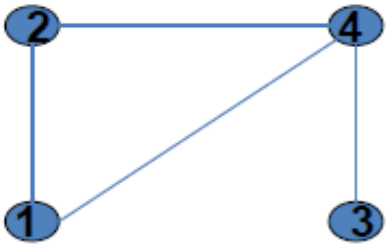
a) Adjacency matrix
b) Random Graph
c) Edge lists
d) All the above

**9. Is the walk from 1 to 3: (1) a path? (2) a cycle?**



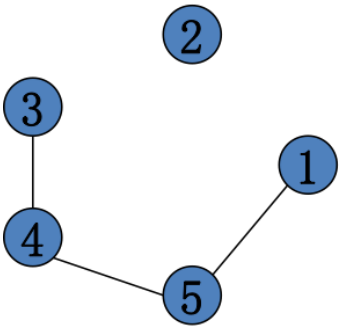

a) Yes, Yes
b) Yes, No
c) No, Yes
d) No, No

**10. What is the closeness centrality of node 3 in the network?**



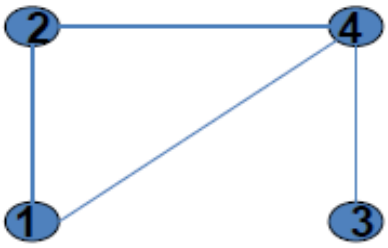

_____

**11. What is normalized degree centrality of node 1 in the network?**




_____

**12. What is normalized betweeness centrality of node 4 in the network?**




_____

**13. What is the diameter of the network from question 11?**

_____

**14. Which network from question 11 or question 12 has higher density?**

_____

**15. What is the mean nodal degree of the network from question 12?**

_____

## Module 2. sample assignments (tentative tasks)

1. Provide 5 examples of composition variables in Network Analysis (may be from your Individual SNA Research Project).

2. Provide 5 examples of structural variables in Network Analysis (may be from your Individual SNA Research Project).

3. Create a data file of your own ego network with minimum 10 nodes. Explain your choice of a file format.

4. Using software of your choice (R, Pajek. UCINET, Gephi) open a data file of your own ego network with minimum 10 nodes. Prepare data for analysis.

5. Using software of your choice (R, Pajek. UCINET, Gephi) visualize and describe your own ego network with minimum 10 nodes. Explain the results.

6. Using software of your choice (R, Pajek. UCINET, Gephi) calculate centrality measures in your own ego network with minimum 10 nodes. Explain the results.

7. Using software of your choice (R, Pajek. UCINET, Gephi) explore the network characteristics of the data file provided. Explain the results.

8. Using software of your choice (R, Pajek. UCINET, Gephi) generate random network with 25 nodes and examine the diameter and average clustering coefficient.

9. Using software of your choice (R, Pajek. UCINET, Gephi) find the most important nodes in the network of the data file provided. Explain the results.

10. Provide an example or develop single name generator (may be from your Individual SNA Research Project).

11. Provide an example or develop resource generator (may be from your Individual SNA Research Project).

12. Provide an example or develop multiple name generator (may be from your Individual SNA Research Project).

13. Provide an example or develop name interpreter (may be from your Individual SNA Research Project).

14. Suggest research design for your Individual SNA Research Project. Explain your choice.

15. Suggest a sampling strategy for your Individual SNA Research Project. Explain your choice.

16. Suggest a method of data collection for your Individual SNA Research Project. Explain your choice.

17. Describe 5 main advantages of secondary data in network analysis.

18. Describe 5 most serious limitations of secondary data in network analysis.

19. Describe 5 main advantages of primary data in network analysis.

20. Describe 5 most serious limitations of primary data in network analysis.

## Assignments & Grading

- Practical class activities and practice tasks:

5 class activities within module 1 (max 2 each) = max 10 / min 6
5 class activities within module 2 (max 2 each) = max 10 / min 6
Overall class activities = max 20 / min 12

- Module control test 1
  (MCT 1 written test during last module class) = max 10 / min 6

- Module control test 2
  (MCT 1 written assignments during last module class) = max 10 / min 6

- Individual SNA Research Project (submitted before deadline & presented during module classes) = max 20 / min 12

***Semester overall = max 60 / min 36***

- ***Final written exam = max 40 / min 24***

**Overall grade = max 100 / min 60**

## 5. Requirements for analytical report of individual network research project

Individual SNA Research Project is required to be submitted in person before deadline in electronic form with the network data file attached and in printed version. It should be presented and discussed during last practical class.
The topic for research is chosen by students according to their interests and should be approved by the course instructor.

Minimal size of the studied network is **25 nodes.**

Required elements of the final analytical report:

- Title & Student's Name
- Problem, Research question & Goal of analysis (what knowledge you intend to obtain and why; relevance of SNA in the research)
- Theoretical approach & Hypotheses
- Measurement (Concepts, Indicators & Operational definitions)
- Cases selection (sampling methods if applicable) & Ethical aspect of research
- Sources of data & Methods of data collection
- Plan & Methods of analysis
- Results of analysis including:
  - Network visualization,
  - Network metrics,
  - Characteristics of nodes,
  - Testing of hypotheses,
  - Meaningful interpretation of results
- Conclusion:
  - main research findings & applications (recommendations if applicable),
  - analysis limitations and further research prospects
- Bibliography
- Instrumentation (Questionnaire, Codebook etc.)
- Data files (electronic)
- Code (if applicable)

A printed version of the report should by single-spaced, Times New Roman 14, all margins 20mm, up to 15 pages (excluding bibliography and instrumentation).

All elements of the report including data files, presentation and answers during discussion will be evaluated.
Maximum mark is 20 points.
12 points are *minimum to pass.*

## 6. Exam questions sample

***List of tentative questions to cover topics during final exam***

1. Origin and History of SNA
2. SNA Application in Social Research.
3. Levels of analysis in SNA
4. Network representation
5. Different types of networks
6. Network Distance Measures
7. Network Structure
8. Network Cohesion
9. Transitivity. Network Density vs. Clustering coefficients
10. Degree Distributions in Networks
11. Random Graphs Models and their application in Network Analysis
12. Theorem on Network Structure
13. Concept of Centrality. Why and how do we use this concept in SNA?
14. Centrality Measures
15. Which Centrality Measures do we use for which purposes?
16. How do we interpret Centrality Measures?
17. Limitations and constraints of various centrality measures in undirected, directed and weighted networks
18. Research design in SNA and other types of sociological research
19. Types of SNA Research design
20. Features of design to study ego networks
21. Cognitive social structure (CSS) research design in SNA
22. Main strategies of sampling in SNA
23. Types, sources and boundaries of network data
24. Measurement in SNA
25. Data collection in SNA and other types of sociological research
26. Data collection methods in SNA
27. Advantages and limitations of primary network data
28. Questionnaires for network data collection
29. Advantages and limitations of secondary network data
30. Network data file formats
31. Software application in SNA
32. Data management in SNA
33. Matrix transposing
34. Visualization of networks
35. Analysis of Network properties
36. Analysis of Cohesive Subgroups
37. Analysis of Affiliations Networks
38. Analysis of Equivalence
39. Analysis of Ego Network data

***List of tentative assignments to cover topics during final exam***

1. Compare closeness centrality of two selected nodes in the network. Interpret the results.

2. Compare betweenness centrality of two selected nodes in the network. Interpret the results.

3. Compare degree centrality of all nodes in the network. Interpret the results.

4. Compare two networks diameter. Interpret the results.

5. Compare two networks density. Interpret the results.

6. Compare two networks mean nodal degree. Interpret the results.

7. Compare two networks global clustering coefficients. Interpret the results.

8. Compare local clustering coefficients of two selected nodes in the network. Interpret the results.

9. Identify types and compare two networks. Interpret the results.

10. Identify formats of network representation. Draw the network graphically.

## 7. Course lectures outline

### Module 1. Analysis of networks in research of social processes

### Topic 1. Theoretical concepts and methodological foundations of SNA

***Lecture learning objectives:***

1. Knowledge of the origin and development of Social Network Analysis
2. Understanding the scope and the power of Social Network Analysis in sociology and other fields
3. Ability to differentiate levels of analysis of social networks
4. Knowledge of fundamental definitions and basic types of networks
5. Ability to identify networks and understand their mathematical representation based on Graph Theory

**Origin and development of network analysis:**

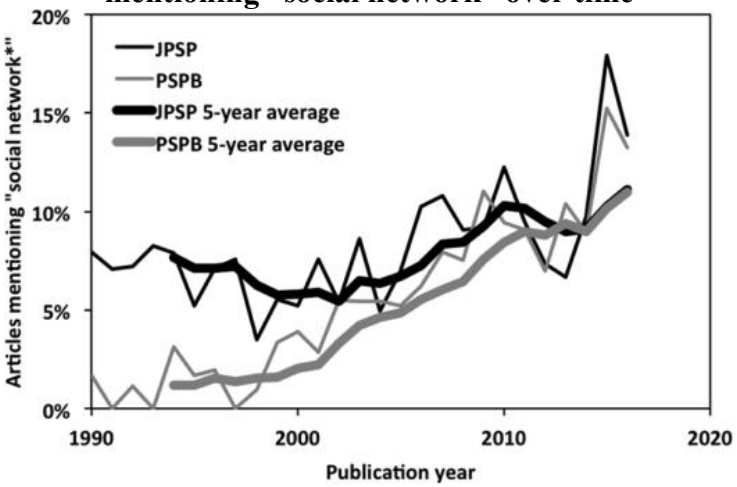
- Origin in anthropology, psychology, sociology, mathematics (Freeman 2004)
- Morgan (1871) family & kinship
- Macfarlane (1883) algebraic model of kinship and graphic images
- Jacob Moreno (1934) - *Sociometry*
- Probabilistic methods: Paul Erdős & Alfréd Rényi (1959) Erdős–Rényi random graph model
- Harrison White - "Harvard Revolution" in Social Networks, Mathematical Sociology (1960s)
- Stanley Milgram (1967) Small-world experiment
- Mark Granovetter (1973) "The Strength of Weak Ties"
- New Economic Sociology (1980s) – embeddedness, relations between actors (individuals and organizations)
- Ronald Burt (1992) social capital, structural holes
- Duncan J. Watts & Steven Strogatz (1998) Watts–Strogatz model
- Albert-László Barabási & Réka Albert (1999) Barabási–Albert model

**Network theory** - the study of graphs as a representation of relations between objects: Sociology, Economics & Finance, Computer Science, Biology, Statistical Physics, etc.

**Social Network Analysis** – theoretical and methodological perspective in Sociology focusing on structural relations between actors.
Modeling social reality (society), relations between actors (individuals, families, households, organizations, states) which mathematically are represented as graphs (networks).

**Social Network Analysis in Ukraine:** Zhulkevska O., Gorbachyk A., Brik T., Dukach Y., Kostuchenko T. and others.

**Percentages of journal articles mentioning "social network" over time**

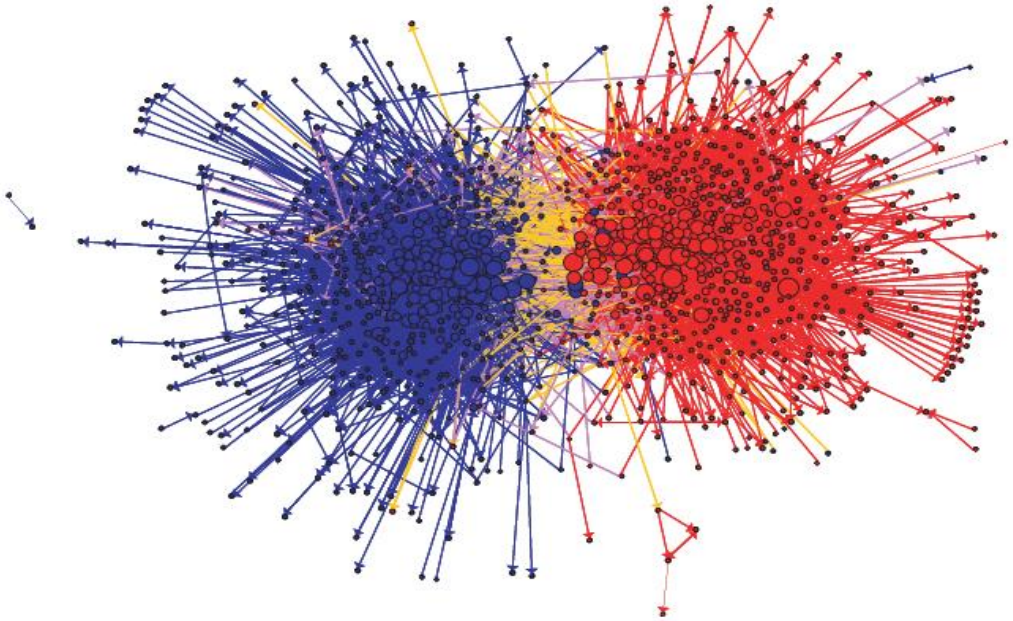



Source: (Clifton, Webster 2017)

## Application of Social Network Analysis in sociology

Focus on relationships not only attribute variables.
Revealing hidden structure.

**Polarization of Political blogs in US (Blue-Democrats, Red-Republicans, Orange-Independent )**



Source: (Lada Adamic 2012 Social Network Analysis)

*Objects of analysis (modeling units)*:

- Actors (positions & roles)
- Dyadic (triadic) analyses (pairs of actors)
- Subgroups (subsets of actors – cliques)
- Whole network characteristics
- Comparing networks

**Scope and capability of Social Network Analysis**

- Are nodes connected through the network?
- How far apart are they?
- Are some nodes more important due to their position in the network?
- How is the network structured? Is it composed of communities?
- How is the network formation happening?
- How does information spread across the network and how is opinion formed?
- How do actors coordinate?
- How durable is the network?

(adapted from Lada Adamic 2012 Social Network Analysis)

**Network representation:**

**Networks** - nodes connected by edges (Graphs).
N={1,…,n} nodes, vertices, agents, actors, players
Connections between nodes: edges, lines, links, ties.
Networks are represented by:
- adjacency matrices,
- edge lists,
- adjacency lists.

Adjacency matrix = edge list = adjacency list = Graphical representation = Network

***Adjacency matrix*** (*sociomatrix)* - "the entries in the matrix indicate whether two nodes are adjacent or not"
***Incidence matrix*** records which edges are incident with which nodes.
(adapted from Wasserman, Faust 1994 Social Network Analysis Methods and Application)

Types of networks:

- One-mode networks - a single set of actors
- Two-mode networks - two sets of actors or
  Affiliation network - one set of actors and one set of events
- *Ego-centered* network – "consists of a focal actor, termed *ego,* as set of alters who have ties to ego
- Directed & undirected networks
- Binary & weighted (valued) networks

## Topic 2. Basic properties and metrics of social networks

***Lecture learning objectives:***

1. Knowledge of fundamental network properties and mathematical measures
2. Ability to describe networks and understand their characteristics
3. Ability to compute network metrics
4. Basic knowledge of Random Graphs and Degree Distribution
5. Understanding the application of Theorem on Network Structure for SNA

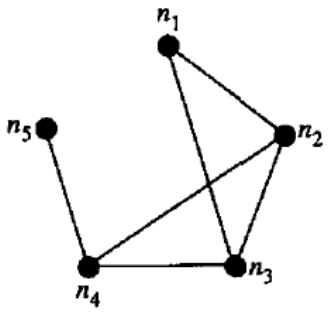
**Network properties**

- Global patterns of networks – path lengths, density, diameter, degree & degree distributions, etc.
- Segregation & Local Patterns –Clustering, Transitivity etc.
- Positions in networks – Neighborhoods, Centrality, Influence etc.

(adapted from Matthew Jackson 2017 Social and Economic Networks: Models and Analysis)

**Reading network**

- ***Walk*** from $i_1$ to $i_K$ is a sequence of nodes ($i_1$, $i_2$,..., $i_K$) and sequence of edges for each k (the edge is incident to both $i_K$ -1 and $i_K$ )
  ***Cycle*** is a walk where $i_1 = i_K$
- ***Path*** is a walk ($i_1$, $i_2$,..., $i_K$) with each node $i_k$ distinct
- ***Geodesic*** - a shortest path between two nodes (minimum number of edges)
- ***Diameter*** is the length of a largest geodesic (largest shortest path), if the network has several components - of largest component
- ***Average Path Length*** **-** the average geodesic - the average number of edges along the shortest paths for all network nodes




**geodesic distances**

$d(1,2) = 1$
$d(1,3) = 1$
$d(1,4) = 2$
$d(1,5) = 3$
$d(2,3) = 1$
$d(2,4) = 1$
$d(2,5) = 2$
$d(3,4) = 1$
$d(3,5) = 2$
$d(4,5) = 1$

$$\text{Diameter of graph} = \max d(i,j) = d(1,5) = 3$$

Source: (Wasserman, Faust 1994 Social Network Analysis Methods and Application)

**Network density**
Real (L) and total possible links that may exist between n nodes ($L_{max}$)
Density = L / $L_{max}$
undirected graph $L_{max}$ = n*(n-1)/2
directed graph $L_{max}$ = n*(n-1)
Density undirected graph = L / (n(n-1)/2)
Density directed graph = L / n(n-1)

Sparse Density in Real World Networks.

**Nodal Degree** $d(n_i)$ – number of incident edges of a node
Out-degree
In-degree

Nodal degree from adjacency matrix

$$\text{Outdegree} = \sum_{j=1}^{n} A_{ij}$$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

example: outdegree for node 3 is 2, which we obtain by summing the number of non-zero entries in the 3rd *row*

$$\sum_{j=1}^{n} A_{3j}$$

$$\text{Indegree} = \sum_{i=1}^{n} A_{ij}$$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

example: the indegree for node 3 is 1, which we obtain by summing the number of non-zero entries in the 3rd *column*

$$\sum_{i=1}^{n} A_{i3}$$

Source: (Lada Adamic 2012 Social Network Analysis)

**Nodal Degree Distribution**
Each node has a certain degree value.
Degree values vary across the network.

→ they form a distribution

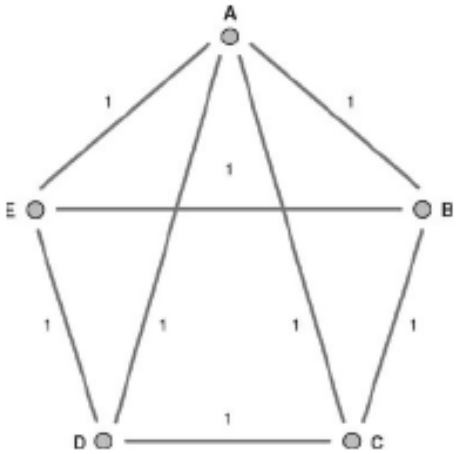
Degree distribution is the probability distribution of nodal degrees over the network
Degree distribution follows statistical laws

**Central tendency measure**

**The mean nodal degree** is a statistic that reports the average degree of the nodes in the graph.
(adapted from Wasserman, Faust 1994 Social Network Analysis Methods and Application)

***Mean nodal degree in undirected networks***

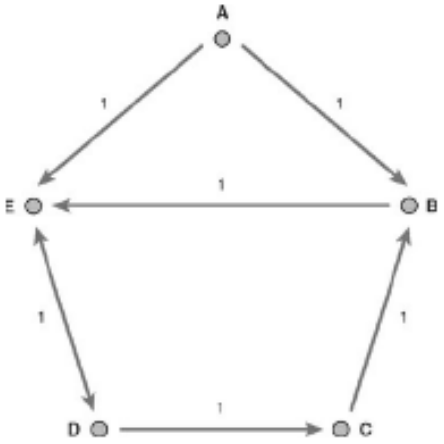
Individual degrees for A, B, C, D, and E are 4, 3, 3, 3, and 3, respectively, the network's mean nodal degree is 16/5 = 3.2




Source: (David Knoke, Song Yang 2008 Social Network Analysis, SAGE)

***Mean nodal degree in directed networks***

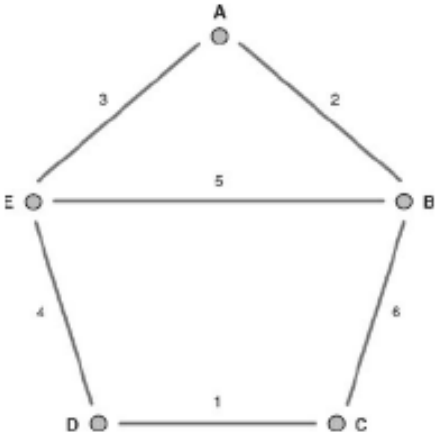
Indegrees for A, B, C, D, and E are 0, 2, 1, 1, and 3, whereas their outdegrees are 2, 1, 1, 2, 1. Both sets aggregate to 7; the network's mean nodal degree is 7/5 = 1.4




Source: (David Knoke, Song Yang 2008 Social Network Analysis, SAGE)

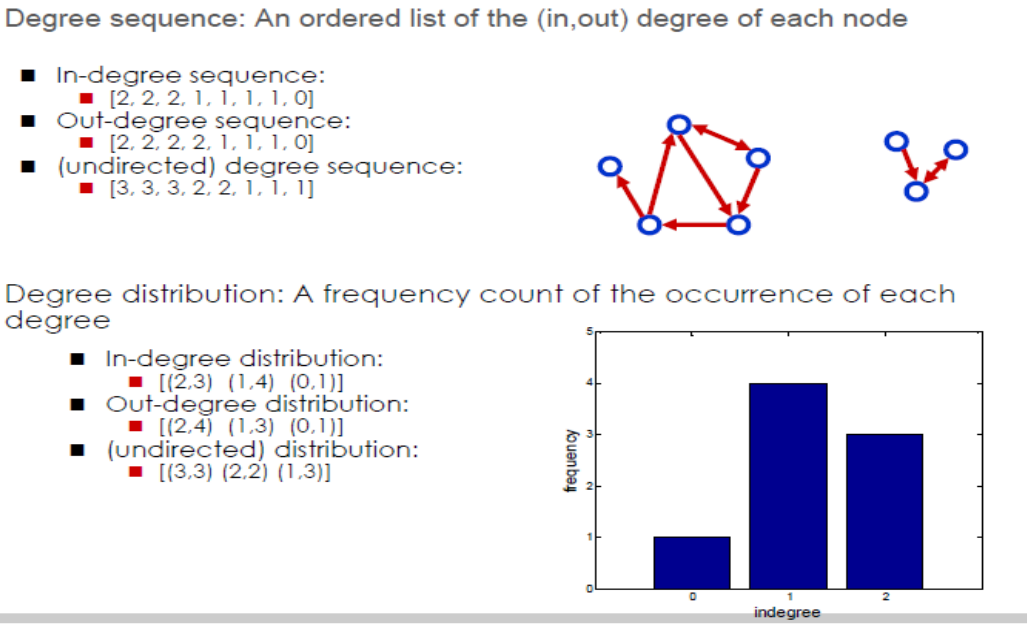
***Mean nodal degree in weighted networks***

**Summation method** – the nodal degrees of A, B, C, D, and E are 5, 13, 7, 5; the network's mean nodal degree is 42/5 = 8.4
**Average method** – average nodal values by number of edges first (e.g. A = 2+3 / 2 = 2.5)




Source: (David Knoke, Song Yang 2008 Social Network Analysis, SAGE)

## Degree Distribution as Network Characteristic




Source: (Lada Adamic 2012 Social Network Analysis)

## Erdos-Renyi Random Graph Models (ER)

Models: *G*(*n*, *p*) or *G*(*n*, *M*) by Paul Erdős & Alfréd Rényi (1959)
Nodes connect at random, probabilities of connection are independent, network is undirected
Parameters:

n = total number of nodes in the graph
p = probability that any two nodes are connected
M = total number of edges in the graph

Degree distribution *G*(*n*, *p*) – *n* nodes where each pair of nodes is connected with probability *p.*
Probability that node has d links is ***binomial distribution*** *B*(*n*, *p*), which is ***Erdos-Renyi Random Graph Model***.

Distribution parameters n $\in$ N0, p $\in$ [0,1], each of n nodes of a random graph is connected (or not) with independent probability p (or 1 − p).

Distribution of degrees of random graphs with large *n* & small *p* is approximately ***Poisson distribution*** P(λ).

Large networks may have degree distributions that approximately follow a ***power law*** $P(d) = d^{-a}$ (Scale Free networks).

Growing random networks are no longer binomial, as probabilities vary with time - probability **m/t** of getting new link each period ("Mathew Effect")

(adapted from Matthew Jackson 2017 Social and Economic Networks)
**Random Graphs vs. Observed Networks**

Observed social networks are more clustered than would be at random.
In reality most of social networks are highly positively-skewed
→ `Fat tails'' compared to random network
When observed networks have `fat tails'' compared to an Erdos-Renyi random network, this refers to the observed network having more high and low degree nodes than predicted at random.
In growing networks new links form with probability proportional to number of links a node already has.

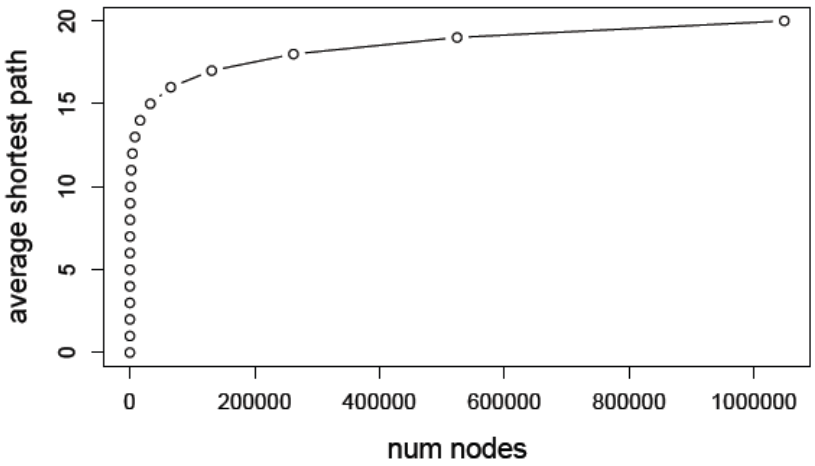
**Theorem on Network Structure**
Under 2 conditions:
1. Links are dense enough so that network is mostly connected (nodal degree is high enough)
2. Network is not too complete (nodal degree is not too high),
then for large n, average path length and diameter are approximately proportional to log(n)/log(d(n))

- *"Six Degrees of Separation" (Frigyes Karinthy 1929)*
- *"Small World" (Milgram 1967)*

This is true for random graphs and observed networks.
(adapted from Matthew Jackson 2017 Social and Economic Networks)

**Small average path length and diameter in random graphs and real world networks**.

**Average shortest path in large networks**




Source: (Lada Adamic 2012 Social Network Analysis)

## Topic 3. Centrality measures in networks

***Lecture learning objectives:***

1. Knowledge of measures for a position in the network
2. Ability to compute centrality measures
3. Ability to interpret centrality measures and describe an actor position in the network

**Positions in networks – Neighborhoods, Centrality, Influence etc.**

Concept of centrality goes beyond of being popular.

Two aspects of Centrality:

1. Centrality is the contribution of the node to the network structure (to which extent an actor is essential for the network = "*structural importance of a node*")
2. Centrality is the advantage of a node by virtue of its position in the network (to which extent an actor benefits from the network)
   → "prominent, or influential, or leaders, or gatekeepers, or as having great autonomy, control, visibility, involvement, prestige, power…"

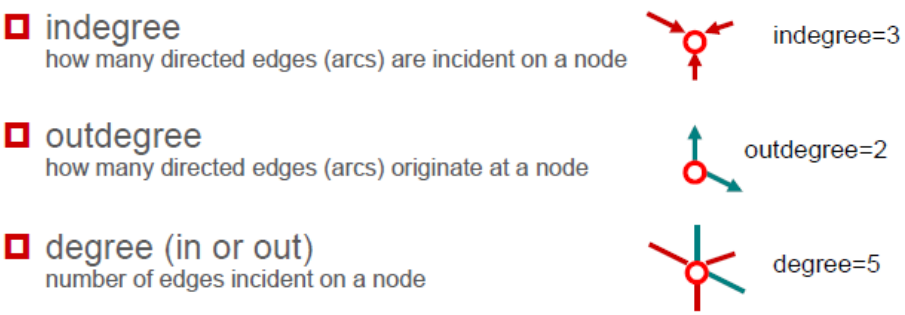
(adapted from Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

A family of concepts – ***different measures of centrality***:

- Degree centrality – connectedness
- Closeness, Decay centrality – ease of reaching other nodes
- Betweenness centrality – role as an intermediary, connector, brokerage
- Eigenvector centrality / Bonacich power centrality (influence, prestige, power – who you know)

(adapted from Matthew Jackson 2017 Social and Economic Networks)

**Degree Centrality** $d(n_i)$ – number of incident edges of a node

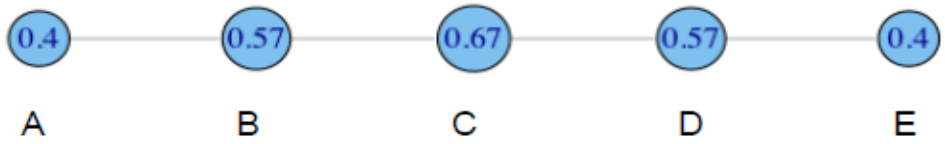



Source: (Lada Adamic 2012 Social Network Analysis)

***Normalized degree centrality*** = $d(n_i) / (n-1)$
(nodal degree ($n_i$) by total possible links)
But not for very large networks.
(adapted from Lada Adamic 2012 Social Network Analysis)

**Closeness Centrality** - the length of the average shortest paths between a node and all other nodes *d*(i,j) in the network, i.e. relative distances to other nodes

$$C_c'(A) = \left[ \frac{\sum_{j=1}^{N} d(A,j)}{N-1} \right]^{-1} = \left[ \frac{1+2+3+4}{4} \right]^{-1} = \left[ \frac{10}{4} \right]^{-1} = 0.4$$




(adapted from Lada Adamic 2012 Social Network Analysis)

**Betweenness (Freeman) Centrality** - How many pairs of actors would have to go through the node in order to reach one another?

$$C_B(i) = \sum_{j<k} g_{jk}(i) / g_{jk}$$

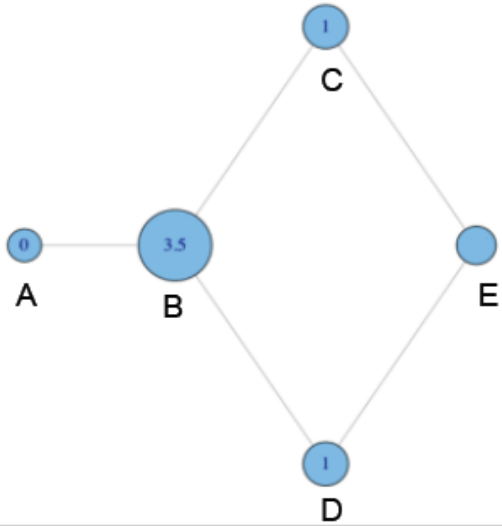
where
$g_{jk}$ = the number of the shortest paths (geodesics) between j and k
$g_{jk}$ *(i)* = the number of the shortest paths (geodesics) between j and k that actor i lies on

Betweenness centrality with alternative paths

C & D are both on the shortest paths
for 2 pairs (AE) & (BE) 2 pairs / 2 paths = 1

B is on the shortest paths for pairs 3 pairs
(AE, AC, AD) 3 pairs / 3 paths = 3
+ 1 pair (CD via B or E ) 1 pair / 2 paths = ½
3 + ½ = 3.5

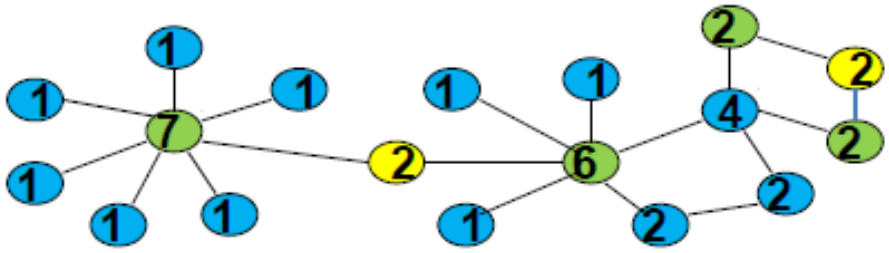
What is the betweenness
centrality of node E?




Source: (Lada Adamic 2012 Social Network Analysis)

Normalized Betweenness centrality = $C_B(i)$ / ((n-1)(n-1)/2)
(adapted from Lada Adamic 2012 Social Network Analysis)

**Measures of Power & Influence**




Source: (Matthew Jackson 2017 Social and Economic Networks)

**Eigenvector centrality** (**eigencentrality)** – an extent to which how central you are depends on how central your neighbors are

$$e_i = \lambda \sum_j x_{ij} e_j$$ where

*e* is the eigenvector centrality score, λ (lambda) is a proportionality constant, $x_{ij}$ adjacency matrix ($x_{ij}$=1 or $x_{ij}$ = 0).
Eigenvector centrality implies that "*a node is only as central as its network*"
(adapted from Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

**Bonacich centrality** is proportional to the sum of centralities not only of the adjacent nodes (your neighbors as reflected by Eigenvector centrality) but to the direct & indirect connections (your neighbors & neighbors of your neighbors) weighted by a length-based parameter β.
β parameter (a length-based weight or attenuation factor) is chosen by the user in order to define whether direct or indirect connections have more weight.
The researcher "*can choose in-between values of β that reflect… conception of how much longer channels of influence matter*"
(adapted from Matthew Jackson 2017 Social and Economic Networks
& Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

**Centrality in Directed and Weighted Networks**
Degree centrality & Bonachich centrality are splint into "*'out' version and 'in' version, reflecting outgoing versus incoming ties*".
Problem of direction and weight of connection between nodes for a meaningful interpretation. A possible solution to make weighted data binary (dichotomization).
(adapted from Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

Centrality measures should be interpreted cautiously in a meaningful way.

## Module 2. Collecting, processing and analysing network data

### Topic 4. Research design, network data collection and processing

***Lecture learning objectives:***

1. Knowledge of peculiarities of research design, sampling, measurement & data collection in Social Network Analysis
2. Ability to design research of social networks
3. Ability to sample network data correctly
4. Ability to find sources, assess limitations of network data and choose a proper method of data collection
5. Ability to measure structural and composition variables and develop a questionnaire for network data collection

**Research (Study) Design in Network analysis** – a specific approach to conduct research and obtain scientific knowledge.
Common research designs: descriptive, explanatory and predictive studies, experimental and quasi-experimental, cross-sectional and longitudinal desigms (examples in Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks). Statistical modeling of network formation and growth, information diffusion and Game theory modeling (examples in Matthew Jackson 2017 Social and Economic Networks).

**Specific types of SNA Research Design**

- ***One-mode whole*** networks design
- ***Two-mode whole*** networks design - two sets of objects:
  - *Cognitive social structure* (CSS) design is focused on "each network member's perceptions of the relationships between all actors in the network" or perceptions of the objects or events from multiple raters
- ***Partial*** networks design (actor-to-group relationships)
  - *Ego-centered (or personal)* network designs "assemble data on relationships involving a focal object (*ego*) and the objects (*alters*) to which it is linked"

(adapted from Models and Methods in Social Network Analysis 2005).

**Choosing SNA Research Design key questions**

- Research goal – descriptive or explanatory
- Network type
- Network size
- Node type - individuals, households, organizations, events, objects, countries
- Type of structural relations
- Hypotheses – how to verify them through network data
- Time dimension - cross-sectional or longitudinal

**Sampling the nodes and relations**

"Sampling of networks has been the most challenging aspect of network research" becauser random (probabilistic) sampling for regular survey does not work.
(adapted from Encyclopedia of Social Networks 2011)

Entire study or Sample?

Entire study of "small but complete populations with a clearly specified social boundary" but population size might be unknown even for small populations (who actually belong to the network).
(adapted from Encyclopedia of Social Networks 2011)

**Strategies of sampling in SNA**

All strategies are aimed at boundaries specification of the network (nominated by research or actor; or induced by researcher: e.g. a certain organization or group):

- "The realist strategy covers social boundaries and memberships based on the subjective perceptions of actors"
- The relational approach is based on social connectedness – selecting key informants who can "provide an accurate and complete list of actors"
- The positional strategy based on characteristics of objects or formal membership criteria and focuses on the attributes of actors who have "similar or equivalent positions in a particular social domain"
- "The event-based approach focuses on actors participating in particular events or activities occurring at specific times and places"

(adapted from Encyclopedia of Social Networks 2011
& Knoke, Yang 2008 Social network analysis)

**Setting Network Boundaries**

| Type of sample | Nominalist/etic (researcher-defined networks) | Realist/emic ('natural' groups) |
|---|---|---|
| Random sample | Random sample of persons matching researcher's criteria. | Do ethnographic pre-study to determine group members, then sample from it. |
| Snowball sample | Interview any qualifying actor with a tie to any actor already selected, up to *K* waves or until quotas or cost limits reached. E.g., ask each person who they inject drugs with, then interview those people. Repeat. | Get starter set of group members. Select all group members with tie to previously selected member. Repeat until few new names appearing. E.g., get self-identified members of gang. Ask them for other members. Repeat. |
| Census | All persons matching researcher criteria. E.g., all members of the Anthropology dept. | Get list of 'members' from somebody in group. E.g., locate gang member, obtain list of members, interview all/adjust on basis of subjective information. |

Source: (Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

**Measuring Network Data**

*Structural* and *composition variables:*

- Composition variables (*attribute* variables) "are measurements of actor attributes"
- Structural variables "measure ties of a specific kind between pairs of actors"

(adapted from Wasserman, Faust 1994 Social Network Analysis Methods and Application)

**Sources of Network Data**

- primary data:
  - survey
  - observation
- secondary sources:
  - existing databases
  - web generated data

**Network Data Collection**

**Survey methods:**

+ allow researchers to define network membership and decide which aspects of relationships to measure"

+ no further data transformation needed

– small or moderate size networks

– rely on the willingness of subjects to self-report their social ties and their memory (access, consent, ethic)

– survey measurement errors (systematic & random)

**Instrumentation for Network Data Collection**

***Name generators*** identify the respondent's alters, i.e. measure structural variables.

Example of 1985 General Social Survey (GSS) Name generator:

"*From time to time, most people discuss important matters with other people. Looking back over the last 6 months—who are the people with whom you discussed matters important to you? 'Do you feel equally close to all these people? (IF NO): Which of these people do you feel especially close to? (PROBE: Anyone else?)"*

***Multiple-generator instrument*** **–** "eliciting persons to whom respondents can turn and on whom they can rely in differing circumstances" (can be long, interview time) → yields larger networks.

***Position generators*** measure personal connections in different social sets.

Respondents are asked if they have any relatives, friends, or acquaintances who hold different social positions (class, occupational, race, ethnic, etc. positions). Position generator data allow construction of indices of network range (e.g.,

number of occupations contacted) and composition (e.g., most prestigious occupation contacted).
(adapted from Models and Methods in Social Network Analysis 2005
& Knoke, Yang 2008 Social network analysis)

Example of Position generator:
*''This is a list of 24 occupations. Please check for each of these occupations the persons you know with such an occupation. So for each occupation you could check multiple persons.''*
*The response categories: ''An acquaintance has this occupation,'' ''A friend has this occupation,'' ''A family member has this occupation' and' ''I know nobody with this occupation.''*
Source: (Verhaeghe 2012 Reliability of Position Generator Measures)

***Resource generator*** measures individual-level social capital as "resources owned by the members of an individual's personal social network, which may become available to the individual". Respondents report if they have access to certain occupations in hierarchical ladders.
Source: (Van der Gaag and Snijders 2005 The Resource Generator)

***Name interpreters*** that obtain information on the alters and their relationships:

- Composition variables (in addition to structural)
- Perception of the network structure (Cognitive social structure CSS - 'Who knows who knows whom?")

(adapted from Models and Methods in Social Network Analysis 2005)

**Observation method:**
Rare alternative to survey wich aloows avoiding large questionnaires and informants' bias but provides limited information.

**Advantages & limitations of secondary data:**
+ Objectivity, fewer measurement errors
+ Large networks (large volume data can be accessed)
+ Relationships of actors who are reluctant to grant interviews
+ Often contain high-quality longitudinal information
+ Relatively inexpensive
- Raw data that was not originally collected for network analysis
- Often time consuming data transformation
- Constrains on research design
- Limited availability, access problem
- Protection of privacy and ethical limitations
(adapted from Models and Methods in Social Network Analysis 2005
& Knoke, Yang 2008 Social network analysis)

## Topic 5. Analysis of networks in programing environments Pajek, Gephi, R or UCINET

***Lecture learning objectives:***

1. Knowledge of peculiarities of network data processing
2. Ability to create a network data file, format, process and transform network data
3. Ability to use various Social Network Analysis software

### Network Data Management

### Software available:

- UCINET http://www.analytictech.com/ucinet/trial.htm
- Gephi https://gephi.org
- Pajek http://mrvar.fdv.uni-lj.si/pajek
- R packages https://cran.r-project.org :
  igraph
  statnet (metapackage, suite of packages including sna, network, ergm)
- Python
- Excel

### Network Data File Types

**Comparison of network data file types**

| | Edge List/Matrix Structure | XML Struture | Edge Weight | Attributes | Visualization Attributes | Attribute Default Value | Hierarchical Graphs | Dynamics |
|---|---|---|---|---|---|---|---|---|
| CSV | ■ | | ■ | | | | | |
| DL Ucinet | ■ | | ■ | ■ | | | | |
| DOT Graphviz | | | ■ | | ■ | | | |
| GDF | | | ■ | ■ | ■ | ■ | | |
| GEXF | | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| GML | | | ■ | ■ | ■ | | | |
| GraphML | | ■ | ■ | ■ | ■ | ■ | ■ | |
| NET Pajek | ■ | | ■ | | ■ | | | |
| TLP Tulip | | | | | | | | |
| VNA Netdraw | | | ■ | ■ | | | | |
| Spreadsheet* | | | ■ | ■ | | | | ■ |

Source: (https://gephi.org)

### How to create Network Data File?

1. In Matrix format
2. In Edge lists format
3. In Adjacency lists format

**Network Data File Format**

1. In a chosen spreadsheet network software format (not recommended)
2. In ASCII format (then import to a chosen network software)
3. In Excel spreadsheet format (then import to a chosen network software)

**Network Data Import & Export**

- UCINET spreadsheet editor can import and export Excel spreadsheets UCINET dataset consists of two separate files that contain header information (filename.##h) and the data lines (filename.##d) and produce datasets in *.DL format
- R and Gephi can import Excel spreadsheets
- UCINET, Pajek, R and Gephi support importing raw ASCII text files
- Pajek can save files in *.net or *.dat ASCII (Text) format; networks & vectors can be exported from Pajek to R or SPSS directly *Tools>SPSS>Send to SPSS*
- Excel files can be saved as ASCII text files

**Network Data in Matrix format**

- Data in Adjacency Matrix format for dense & small networks
- Network may be binary, directed or weighted
- Adjacency Matrix must be imported first
- Attribute Matrix is to be imported separately
- All nodes must be in the same order in Adjacency Matrix & Attribute Matrix

*Example of Network data file in UCINET *DL Matrix format*

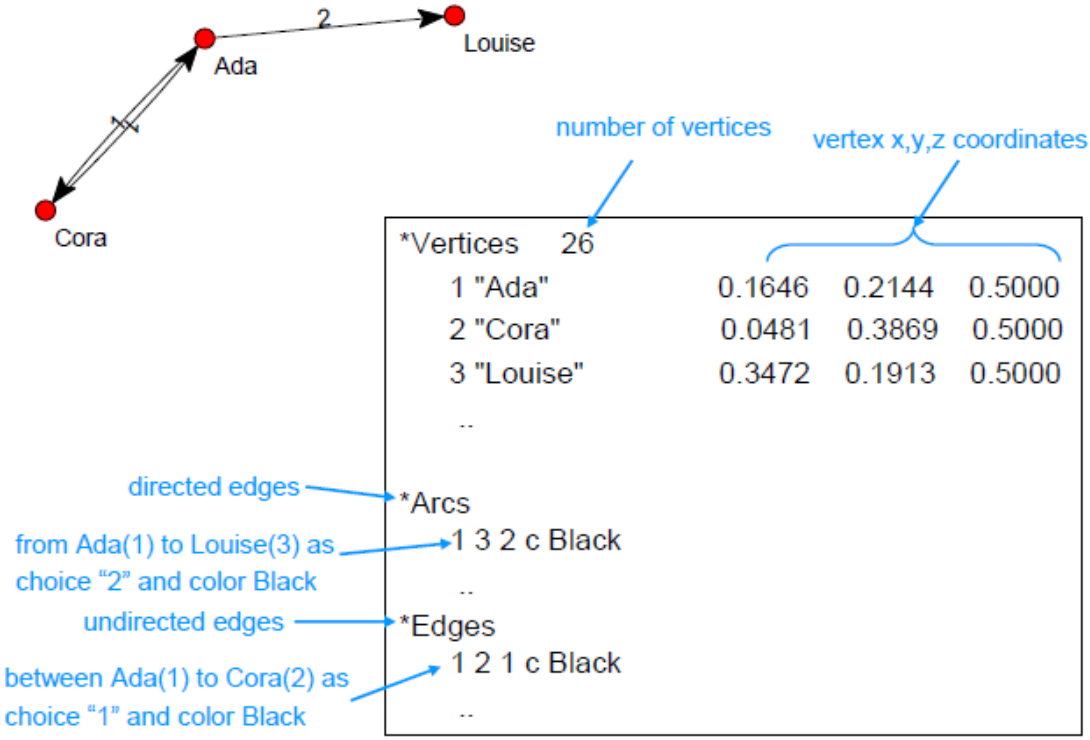
```
DL
N=6 NM=2
FORMAT = FULLMATRIX DIAGONAL PRESENT
ROW LABELS:
ACCIAIUOL
ALBIZZI
BARBADORI
BISCHERI
CASTELLAN
MEDICI
COLUMN LABELS:
ACCIAIUOL
ALBIZZI
BARBADORI
BISCHERI
CASTELLAN
MEDICI
LEVEL LABELS:
PADGM
PADGB
DATA:
 0 0 0 0 0 0
 0 0 0 0 0 1
 0 0 0 0 1 0
 0 0 0 0 0 0
 0 0 1 0 0 0
 0 1 0 0 0 0
```

Source: (Matthew Jackson 2017 Social and Economic Networks)

**Network Data in Edge list format**

- Data in Edge list format for sparse & bigger networks
- Network may be binary, directed or weighted
- Attributes may be in the same file
- Matrix can be read as edge list (UCINET *format=edgelist1* command for one-mode network; as well as in R igraph or statnet)

*Example of Network data file in Pajek *.net Edge list format*




Source: (Lada Adamic 2012 Social Network Analysis)

**Digitally Generated Network Data**

Web data are collected using scrapers and Web crawler (spiders)

- Scrapers are automated computer scripts that take a web page and transform its content making it useful as data
- Spiders are a special class of scraper that follow links and collect information along the way returning a set of node-node pairs between a 'seed set' and the pages to which they are linked.
- Special software tools and coding (e.g. Scrapy written in Python; Java; R)

(adapted from Analyzing Social Networks via the Internet 2008)

**Network Data Processing**

Preparing Data for Analysis:

Cleaning, Matching attributes and networks, Missing Values, Recoding, Selecting, Combining, Transposing, etc.

UCINET data transformation commands:

*Transform>Recode* (re-assign values)

*Transform>Reverse*
*Transform>Dichotomize* (turning valued data into binary data for further analysis)
*Transform>Collapse* (combining rows and/or columns)
*Transform>Symmetrize* (to turn "directed" or "asymmetric" network data into "un-directed" or "symmetric" data to apply methods for symmetric data or for data cleaning; different options)
*Transform>Normalize* (for comparison of valued data or networks)
*Data>Attribute* (turning relations into attributes)

***Selecting network sub-sets*** of the data (to delete a node or nodes, rows, columns, e.g. not connected nodes; extract sub-graphs, etc.)
UCINET data selection commands:
*Data>Extract*
*Data>Subgraphs from partitions*
*Data>Remove isolates*
*Data>Remove pendants*
*Data>Egonet* (extraction of a particular actor ego-network)
*Data>Extract main component* (extraction of the largest component of the network)
*Data>Join* (to combine separate sets of data into a new data set)
(adapted from Hanneman, Riddl 2005 Introduction to social network method)

***Matrix Transposing*** - interchanging matrix rows with its columns (including reversing the direction of the arcs) might be needed for meaningful interpretation of relations & imputing missing data for symmetrical networks.

UCINET command:
*Data>Transpose*

This operation also starts analysis of data because “the correlation between an adjacency matrix and the transpose of that matrix is a measure of the degree of reciprocity of ties. Reciprocity of ties can be a very important property of a social structure because it relates to both the balance and to the degree and form of hierarchy in a network”.
(adapted from Hanneman, Riddl 2005 Introduction to social network method
& Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

## Topic 6. Methods of analysis of network data

***Lecture learning objectives:***

1. Knowledge of basic methods of Social Network Analysis
2. Ability to choose appropriate methods of Social Network Analysis
3. Ability to visualize networks correctly
4. Ability to analyze actors positions and cohesive subgroups in networks
5. Ability to analyze the properties and compare networks
6. Ability to analyze two-mode (affiliation) and partial networks
7. Ability to interpret and present the results of analysis of network data

**Scope and Levels of Analysis of network data.**

1. Visualization - qualitative mapping of a network for rapid pattern recognition
2. Mathematical description – objective network metrics
3. Statistical models (inference & hypotheses testing, relationships, prediction)

**Modeling units (**based onTopic 1**):**

- Actors (positions & roles)
- Dyadic (triadic) analyses (pairs of actors)
- Subgroups (subsets of actors – clusters, cliques)
- Whole network (structure & properties)
- Partial networks (e.g. components, Ego networks)
- Comparing different networks (matching, relating, longitudinal comparisons)

**Steps in Analysis of network data:**

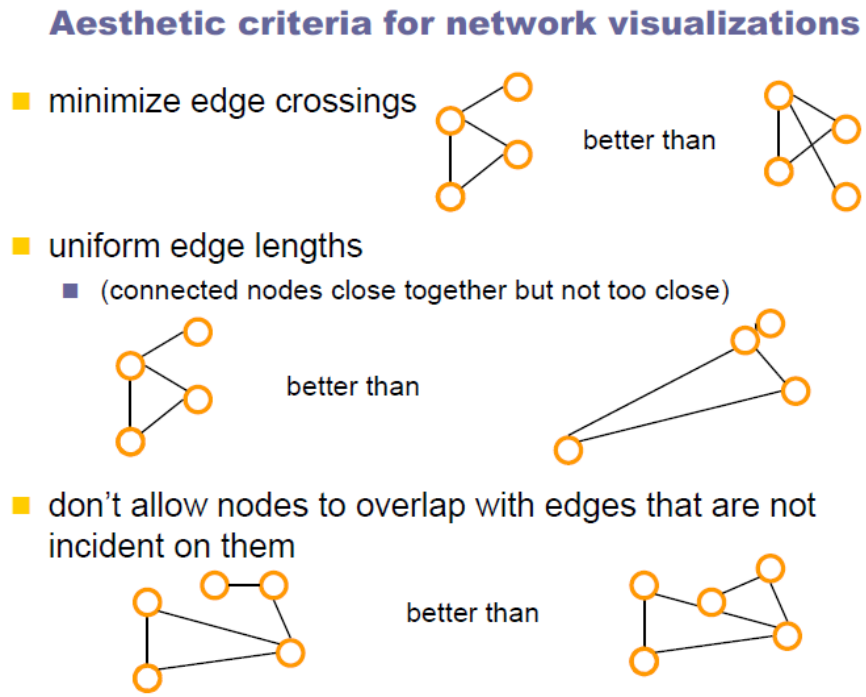
- First step in network analysis is visualization (an introductory but necessary step)
- Considering the network as a whole: density, connectedness, centralization and core-periphery indices
- Network structure: subgroups and community detection
- Considering the key players in the network: centrality (Degree centrality, Closeness centrality Betweenness centrality, Eigenvector centrality, Bonachich centrality and other)
- Relational-Level Analyses - attributes of network members: homophily and assortativity
- Hypotheses testing on attributes & relations
- Modelling relation between network structure and actors' behavior

(adapted from Matthew Jackson 2017 Social and Economic Networks & Analyzing Social Networks via the Internet 2008)

**Network Visualization**

- Use of color
  - nodes may be colored by outdegree
  - edges may be colored by degree of endpoints
  - nodes may be colored by nodal attributes
- Use of meaningful coordinates
  - distance from the most central node
  - position along ring denotes geographical latitude
- Use of different sizes and shapes
  - nodes sized by degree
  - node shape
  - edge thicknes (by value of relation in weighted networks)
- Reduce the number of nodes and edges in large notworks
  - introduce thresholds (e.g. only edges with weight > y; only nodes with degree > z)
  - collapse nodes into clusters (show multiple nodes as a single node; display connections between clusters instead of nodes)

*"Highlighting node attributes through shape and color enhances understanding".*
*"Cool looking visualizations are not always most informative".*
(adapted from Lada Adamic 2012 Social Network Analysis)




Source: (Lada Adamic 2012 Social Network Analysis)

**Whole Network Analysis (**based onTopic 2**)**
Measures of transitivity, reciprocity and clustering.
Centralization, core-periphery shape, cohesion and reciprocal ties.

**Network structure: cohesive subgroups and community detection**
Observed social networks are more clustered than random graphs because of trasitivity.
***Transitivity***: if A is connected to B and B is connected to C, what is the probability that A is connected to C?
"A relation is *transitive* if whenever *iRj* and *jRk*, then *iRk*, for all *i*, j, and *k"*
(adapted from Wasserman, Faust 1994 Social Network Analysis Methods and Application)

***Global (overall, standard) clustering coefficient***
C(g) = 3*closed triads / all triads

***Local clustering coefficient***
$C_{i\ undirected}$ = $n_i$ transitive pairs / ($n_i$ ($n_i$ -1)/2)
$C_{i\ directed}$ = $n_i$ transitive pairs / ($n_i$ ($n_i$ -1))

***Average (Weighted) Local clustering coefficient*** (Watts & Strogatz 1998)

$$C = \frac{1}{n}\sum_i C_i$$

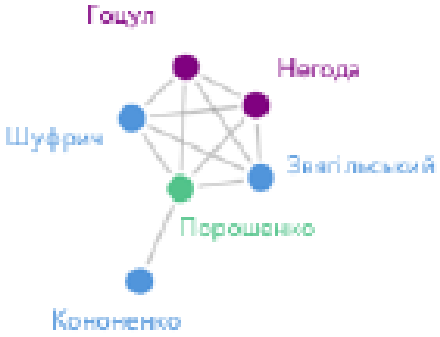
Cohesive subgroups: ***N-cliques***
$K_n$ is the complete graph (clique) with K vertices

- each vertex is connected to every other vertex
- there are n*(n-1)/2 undirected edges

(adapted from Lada Adamic 2012 Social Network Analysis)

***Clique example***

***5-clique + 1 node***
President Poroshenko (green node) formed a 5-clique through ownership of shares in Prominvetbank with MPs Nestor Shufrych and Yukhym Zviahilskyi, as well as with First Deputy Minister of Youth and Sports Ihor Hotsul and First Deputy Minister of Regional Development, Construction and Housing Viacheslav Nehoda




Source: (VoxUkraine 2018 Property – connecting politicians)

**Key actors & centrality measures (**based onTopic 3**)**
UCINET commands:
*Network>Centrality>Degree*
*Network>Centrality>Closeness*
*Network>Centrality>Eigenvector*

*Network>Centrality>Influence*
*Network>Centrality>Betweenness>Nodes*
*Network>Centrality>Betweenness>Hierarchical Reduction*
(adapted from Hanneman, Riddl 2005 Introduction to social network method)

**Ego networks**

Analysis of ego networks is focused either on ***social capita***l or ***social homogeneity***:

1. Actor's social ties enable access to network resources and support.
2. Social homogeneity aspect is about "how ego's ties determine ego's attitudes and behavior"
   - Estimtion of Ego–alter similarity on attributes.
   - Ego network structural shape measures (density, clusters, structural holes)
   - Key actors in a partial network (centrality, brokerage)

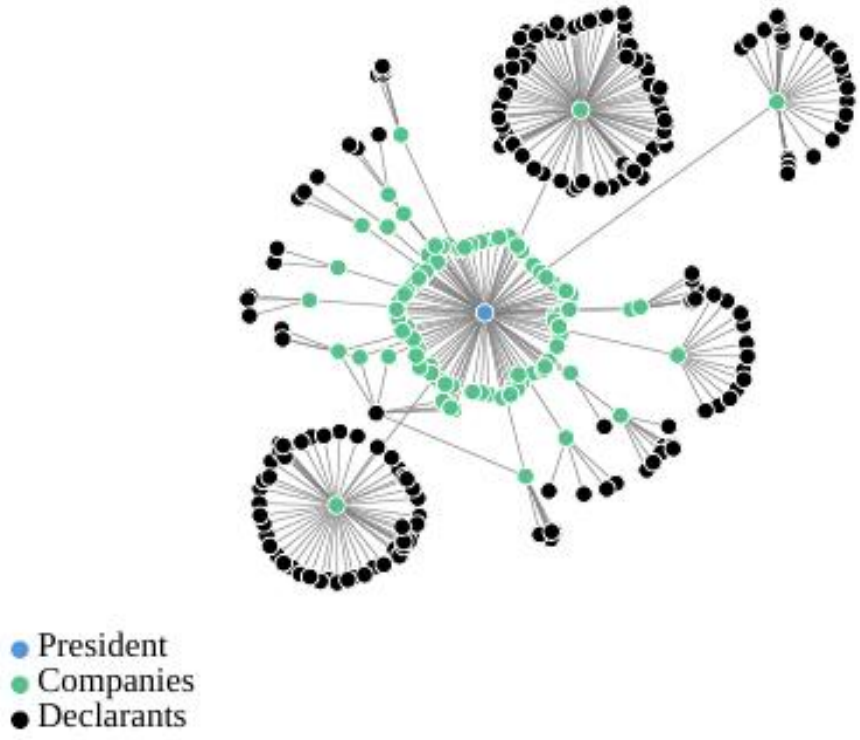
(adapted from Borgatti S., Everett M., Johnson J. 2013 Analyzing Social Networks)

UCINET commands:

*Network>Ego networks>Density*
*Network>Ego Networks>Structural Holes* (position of each actor in their neighborhood)
*Network>Ego Networks>Brokerage* (how ego connects groups)

(adapted from Hanneman, Riddl 2005 Introduction to social network method)

**President Poroshenko's (blue node) ego-network through ownership of shares**




Source: (VoxUkraine 2018 Property – connecting politicians)

## 8. Course literature and resources

***Required textbooks:***


1. Borgatti S., Everett M., Johnson J. Analyzing Social Networks. - London: SAGE, 2013.
2. Luke D. A Users Guide to Network Analysis in R. - Springer, 2015.
3. Wasserman S., Faust K. Social Network Analysis. - Cambridge: Cambridge University Press, 1994.
4. de Nooy W., Mrvar A., Batagelj V. Exploratory social network analysis with Pajek. 2nd ed. - Cambridge: Cambridge University Press, 2011.
5. Gephi Tutorials: Learn how to use Gephi. URL: https://gephi.org/users/


***Optional literature:***


1. Barabási A. L. Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades. - Penguin, 2010.
2. Granovetter The Strength of Weak Ties. *American journal of sociology*, 1973. Volume 78 Number 6
3. Easley D., Kleinberg J. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. - Cambridge: Cambridge University Press, 2010.
4. Encyclopedia of social networks / Barnett George A., ed. - Sage Publications, 2011.
5. Knoke D., Yang, S. Social network analysis (Vol. 154). London: SAGE, 2008
6. Models and Methods in Social Network Analysis. - Cambridge: Cambridge University Press, 2005
7. Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American journal of sociology*, *98*(6), 1259-1319.
8. Scott J. Social network analysis. 4th edition. - London: Sage, 2017.
9. UCINET 6 for Windows USER'S GUIDE. 2002.
10. Горбачик А., Жулькевська О. Мережевий підхід до вивчення структури українського парламенту // Соціологія: теорія, методи, маркетинг. - 2006. - № 3. - С. 161-181.
11. Дукач Ю. Дослідження суспільних рухів за допомогою методу аналізу соціальних мереж // Соціологія: теорія, методи, маркетинг. - 2015. - № 3. - С. 108-124.
12. Костюченко Т., Нагорняк К. Структура зв'язків між депутатами Верховної Ради України 6-го та 7-го скликань: порівняння мереж формування політики // Наукові записки НаУКМА. Соціологічні науки. Том 148. - 2013. - С. 38-44.
13. Сальнікова С. Математичне моделювання соціальних мереж. Навч. посібник для студ. спец. «Соціологія». - Луцьк, 2018. - 120 с.

14. Сальнікова С. Онлайн-дослідження соціальної мережі соціологічних журналів України // Соціологія: теорія, методи, маркетинг. - 2018. - № 4. - С. 135-156.
15. Савельєв Ю.Б. Метод мережевого аналізу у дослідженні соціальних спільнот і актуальні проблеми статистичного моделювання // Проблеми розвитку соціологічної теорії: матеріали XVI Міжнар. наук. конф. «Проблеми розвитку соціологічної теорії: Спільноти: суспільна уява і практики конструювання». - К.: Логос, 2019. - С. 206-208.

**Resources:**

1. Gephi https://gephi.org
2. Pajek http://mrvar.fdv.uni-lj.si/pajek
3. R packages https://cran.r-project.org
4. Freeman Linton (2004). The development of social network analysis. *A Study in the Sociology of Science* https://www.researchgate.net/publication/239228599_The_Development_of_Social_Network_Analysis
5. Hanneman R., Riddle M. Introduction to Social Network Methods https://www.researchgate.net/profile/Robert_Hanneman/publication/235737492_Introduction_to_Social_Network_Methods/links/0deec52261e1577e6c000000/Introduction-to-Social-Network-Methods.pdf
6. Lada Adamic 2012 Social Network Analysis, University of Michigan https://open.umich.edu/find/open-educational-resources/information/si-508-networks-theory-application
7. Matthew Jackson 2017 Social and Economic Networks: Models and Analysis, Stanford University https://www.coursera.org/learn/social-economic-networks
8. VoxUkraine 2018 Property – connecting politicians: a network analysis of Ukrainian top officials' declarations https://voxukraine.org/longreads/declarations-graph/index-en.html
9. UCINET free trial version valid for 60 days http://www.analytictech.com/ucinet/trial.htm

**Навчальне видання**

(англійсько.мовою)

Юрій Борисович Савельєв
доктор соціологічних наук, доцент,
доцент кафедри методології та методів соціологічних досліджень
yursave@knu.ua
Web of Science ResearcherID www.researcherid.com/rid/D-3817-2016
orcid.org/0000-0003-2023-4472
Scopus Author ID 57189896685

# Аналіз соціальних мереж

Навчально-методичний комплекс

для студентів спеціальності
054 «соціологія»
освітній рівень «магістр»

Київський національний університет імені Тараса Шевченка
Факультет соціології

Підписано до друку 7.05.2020.
Ум. друк. арк. 3,2.