A Gentle Introduction to Computational Social Network Analysis

Track 3: Tools for Social Network Analysis

Jaderick P. Pabico
Research Collaboratory for Advanced Intelligent Systems
Institute of Computer Science
University of the Philippines Los Baños
College 4031, Laguna

3rd Philipine Health Care and Social Media Summit (#HealthXPH 2017)
Marco Polo Hotel, Cebu City, 25 April 2017

Acknowledgements

- DOST-PCHRD for the invitation

Acknowledgements

- DOST-PCHRD for the invitation
- #HealthXPH for positioning with technology public health service, research, and instruction in the Philippines

Acknowledgements

- DOST-PCHRD for the invitation
- #HealthXPH for positioning with technology public health service, research, and instruction in the Philippines
- Sir Rick Jason Obrero for using his network so that we can share what we know a little about computational social networks

Objectives

- Main: To introduce to the workshop participants the computational aspects of social network analysis.

Objectives

- Main: To introduce to the workshop participants the computational aspects of social network analysis.
- Specific:
  - To describe the computational data structures of social networks

Objectives

- Main: To introduce to the workshop participants the computational aspects of social network analysis.
- Specific:
  - To describe the computational data structures of social networks
  - To discuss some quantitative metrics of social networks

Objectives

- Main: To introduce to the workshop participants the computational aspects of social network analysis.
- Specific:
  - To describe the computational data structures of social networks
  - To discuss some quantitative metrics of social networks
  - To introduce a free software system for social network analysis

Workshop Outline

- Computational data structures
  - Graphs and Sociograms
  - Matrices
  - Linked list

Workshop Outline

- Computational data structures
  - Graphs and Sociograms
  - Matrices
  - Linked list
- Network metrics
  - Basic Metrics
  - Network Centralities
  - Classifying Nodes

Workshop Outline

- Computational data structures
  - Graphs and Sociograms
  - Matrices
  - Linked list
- Network metrics
  - Basic Metrics
  - Network Centralities
  - Classifying Nodes

Free software system

Pajek*

*Batagelj V and Mrvar A. 2003. *Pajek - Analysis and Visualization of Large Networks*. In **Graph Drawing Software**, Springer pp. 77-103,

Data Structures

What
is this?!

Ay, ano ga
are?!

Atwhay
isway isthay?!

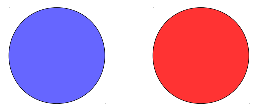
How to represent social network data?

- Visual: Graphs and Sociograms*

* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

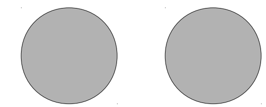
How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices



* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside*.

How to represent social network data?

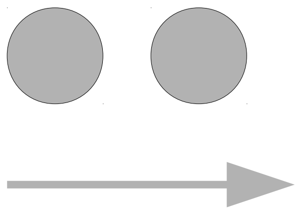
- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices 
  - Relations, Ties, Links, Edges, Arcs 

* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices
  - Relations, Ties, Links, Edges, Arcs



  - Example 1** "*perception of close friendship*"

* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices
  - Relations, Ties, Links, Edges, Arcs
  - Example 1** “*perception of close friendship*”

Blue nodes are males

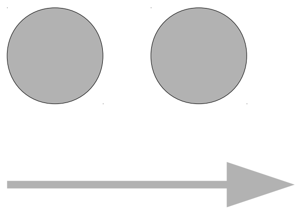
Red nodes are females

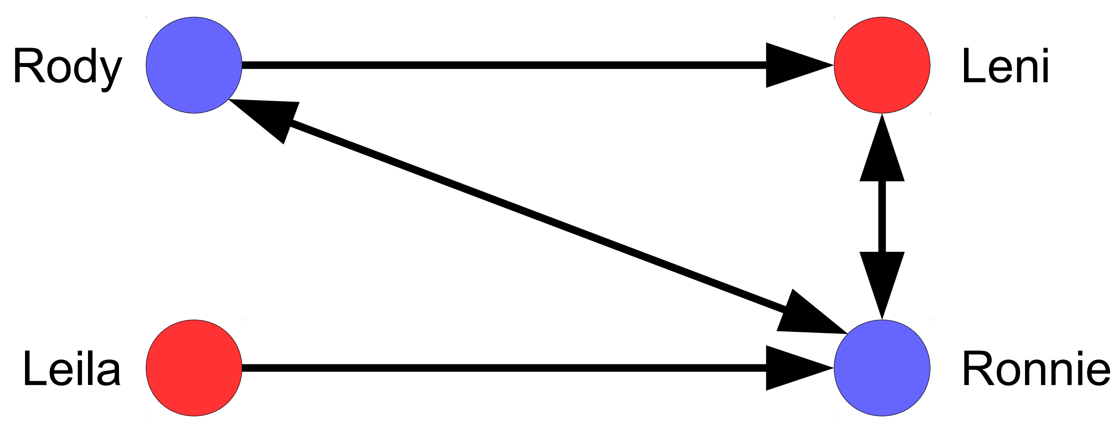
* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices
  - Relations, Ties, Links, Edges, Arcs
  - Example 1** “*perception of close friendship*”

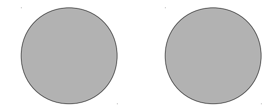


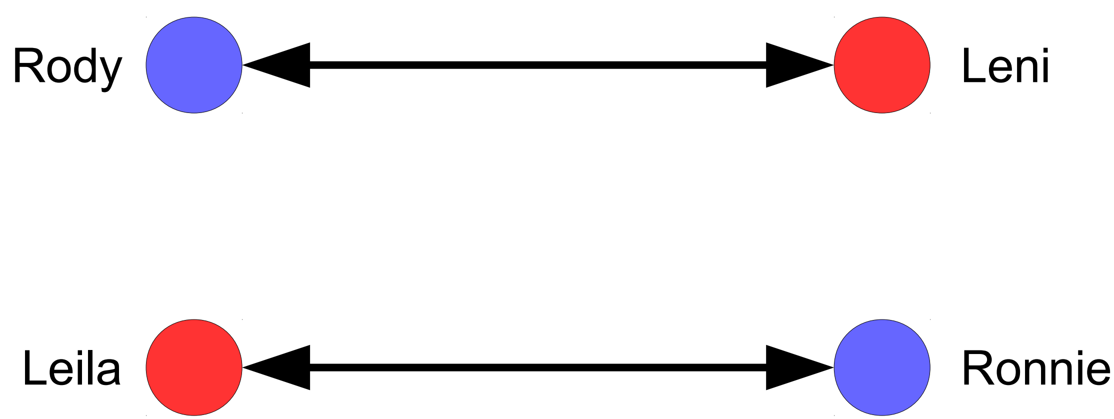


* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside*.

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices 
  - Relations, Ties, Links, Edges, Arcs 
  - Example 2** “*spouse – reciprocated relations*”

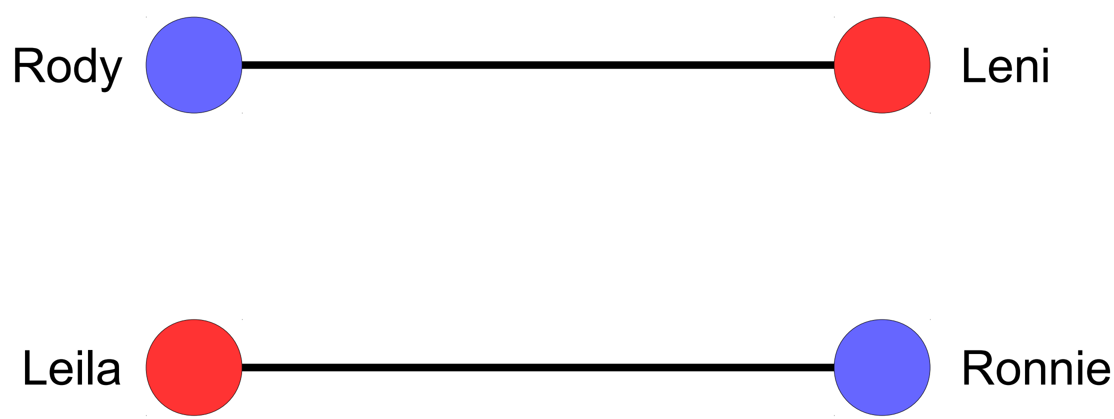


* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside*.

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices
  - Relations, Ties, Links, Edges, Arcs
  - Example 2** “*spouse – reciprocated relations*”



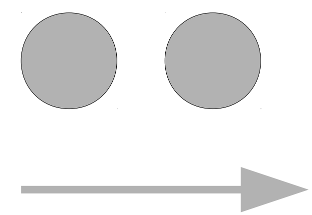
Let us just remove the arrows for clarity.

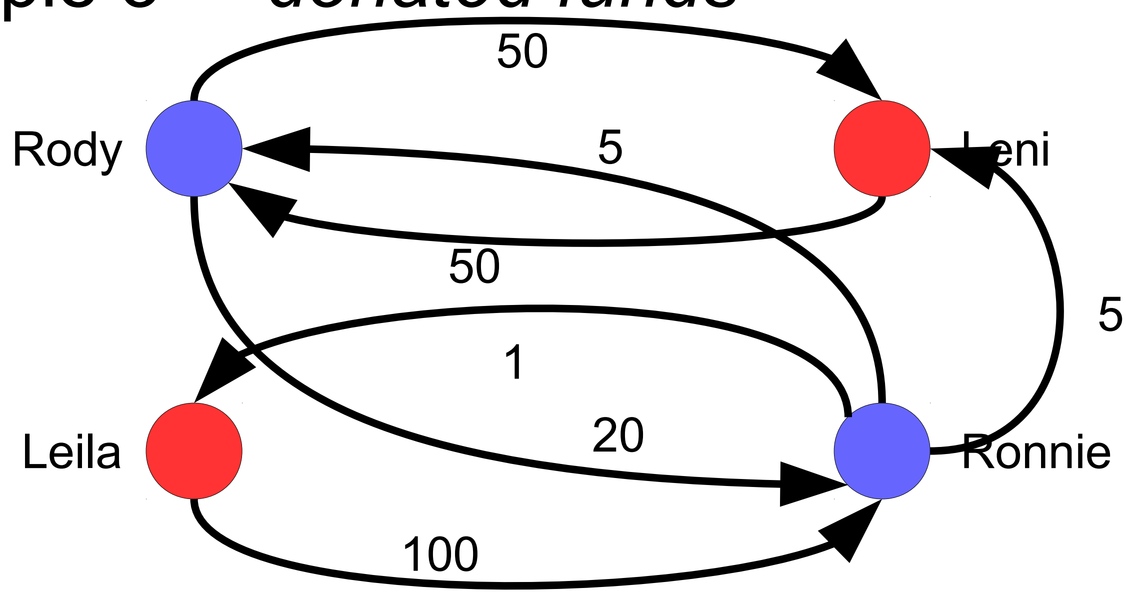
* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices
  - Relations, Ties, Links, Edges, Arcs
  - Example 3** “*donated funds*”



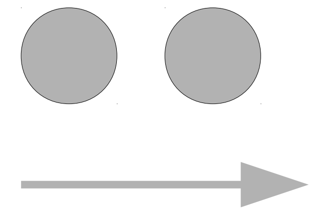


* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

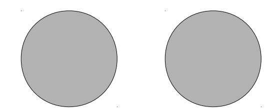
- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices
  - Relations, Ties, Links, Edges, Arcs
  - Example 1 "*perception of close friendship*"
    - Directed network



* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

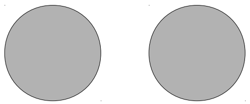
- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices 
  - Relations, Ties, Links, Edges, Arcs 
  - Example 1 "*perception of close friendship*"
    - Directed network
  - Example 2 "*spouse – reciprocated relations*"
    - Undirected network

* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

- Visual: Graphs and Sociograms*
  - Actors, Entities, Nodes, Vertices



  - Relations, Ties, Links, Edges, Arcs

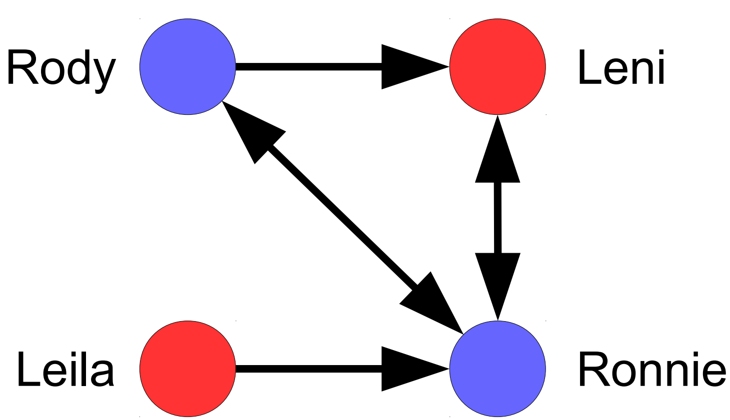
  - Example 1 "*perception of close friendship*"
    - Directed network
  - Example 2 "*spouse – reciprocated relations*"
    - Undirected network
  - Example 3 "donated funds"
    - Directed, weighted network

* Discussion adopted from *Hanneman RA and Riddle M. 2005.* ***Introduction to Social Network Methods****. Riverside, CA: University of California, Riverside.*

** The names used in this example were intentionally made factual. Any similarity to fiction is just a result of probability.

How to represent social network data?

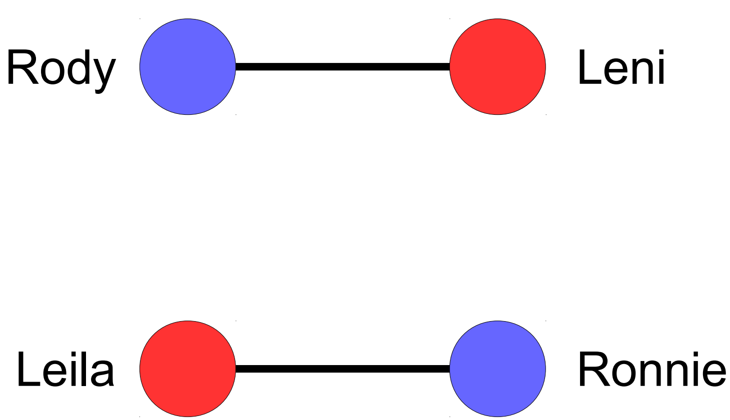
- Matrices
  - Adjacency Matrix, A
    where $A_{j,k} = 1$ if entity j has a relation with entity k, otherwise $A_{j,k} = 0$.
  - Example, directed network



$$A = \begin{array}{c} \begin{array}{cccc} \text{Rody} & \text{Leni} & \text{Ronnie} & \text{Leila} \end{array} \\ \left(\begin{array}{cccc} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \end{array} \begin{array}{l} \\ \text{Rody} \\ \text{Leni} \\ \text{Ronnie} \\ \text{Leila} \end{array}$$

How to represent social network data?

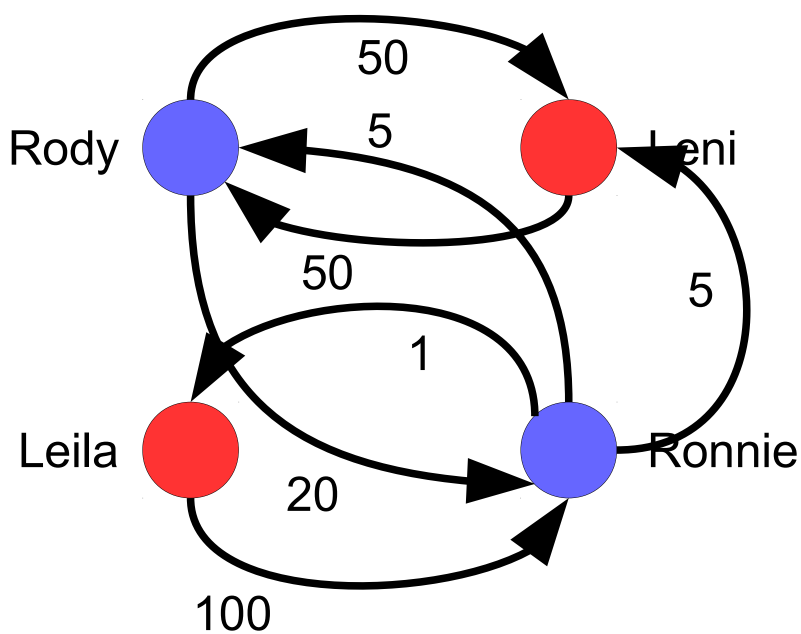
- Matrices
  - Adjacency Matrix, A
    where $A_{j,k} = 1$ if entity j has a relation with entity k, otherwise $A_{j,k} = 0$.
  - Example, undirected network



$$A = \begin{array}{c c} & \begin{array}{cccc} \text{Rody} & \text{Leni} & \text{Ronnie} & \text{Leila} \end{array} \\ & \left(\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array} \right) \begin{array}{l} \text{Rody} \\ \text{Leni} \\ \text{Ronnie} \\ \text{Leila} \end{array} \end{array}$$

How to represent social network data?

- Matrices
  - Weighted Adjacency Matrix, W
    where $W_{j,k} = w$ if entity j has a weighted relation w with entity k, otherwise $W_{j,k} = 0$.
  - Example, weighted directed network



$$W = \begin{array}{cc} & \begin{array}{cccc} \text{Rody} & \text{Leni} & \text{Ronnie} & \text{Leila} \end{array} \\ & \left(\begin{array}{cccc} 0 & 50 & 20 & 0 \\ 50 & 0 & 0 & 0 \\ 5 & 5 & 0 & 1 \\ 0 & 0 & 100 & 0 \end{array} \right) \begin{array}{l} \text{Rody} \\ \text{Leni} \\ \text{Ronnie} \\ \text{Leila} \end{array} \end{array}$$

How to represent social network data?

- Other Matrices*
  - Degree matrix, *D*
  - Normalized adjacency matrix, *N*
  - Laplacian matrix, *L*
  - Normalized Laplacian matrix, *Z*
  - Stochastic adjacency matrix, *P*
  - Signless Laplacian, *K*
  - and many more

Unfortunately, they are “boring” things to talk about.

* *Brouwer AE and Haemers WH. 2011.* ***Spectra of Graphs****. Springer.*

How to represent social network data?

- Lists
  - List of vertices, V
  - List of edges, E

Efficient representation for computation

How to represent social network data?

- Lists
  - List of vertices, V
  - List of edges, E

  Efficient representation for computation

- Example 1: "*perception of close friendship*"
  - V = { Rody, Leni, Ronnie, Leila }
  - E = { (Rody, Leni), (Rody, Ronnie), (Leni, Ronnie), (Ronnie, Rody), (Ronnie, Leni), (Leila, Ronnie) }

How to represent social network data?

- Lists
  - List of vertices, *V*
  - List of edges, *E*

Efficient representation for computation

- Example 2: "*spouse – reciprocated relations*"
  - *V* = { Rody, Leni, Ronnie, Leila }
  - *E* = { (Rody, Leni), (Rody, Ronnie) }

How to represent social network data?

- Lists
  - List of vertices, V
  - List of edges, E

  Efficient representation for computation

- Example 2: "*donated funds*"
  - V = { Rody, Leni, Ronnie, Leila }
  - E = { (Rody, Leni, 50), (Rody, Ronnie, 20),
    (Leni, Rody, 50), (Ronnie, Rody, 5),
    (Ronnie, Lenie, 5), (Ronnie, Leila, 1),
    (Lenie, Ronnie, 100) }

Network Metrics

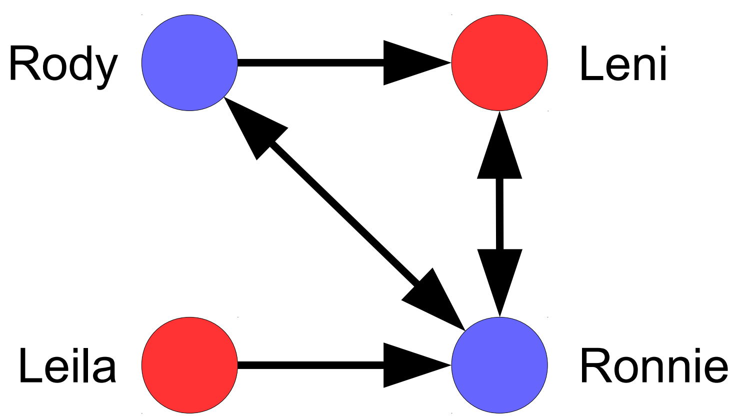
What
is this?!

Ay, ano ga
are?!

Atwhay
isway isthay?!

Network Metrics

- Basic Metrics
  - **Network size**, n – Total number of nodes
  - **Network volume**, m – Total number of edges
  - **Network weight**, w – Sum of absolute edge weights
  - **Average degree** (or Network density), $d = 2m/n$



$n = 4$

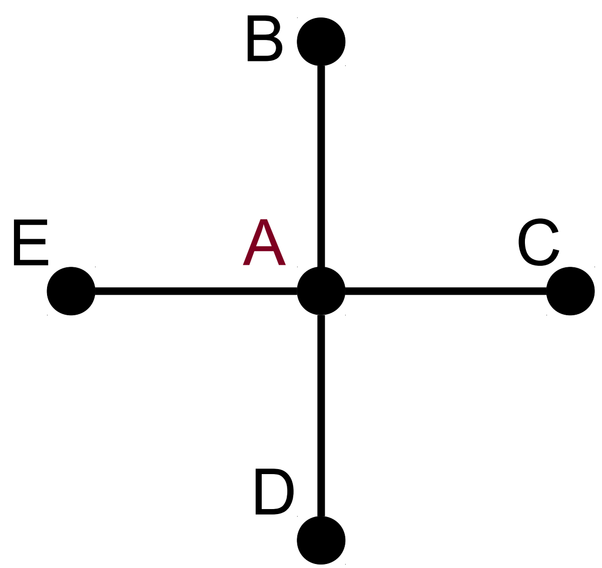
$m = 6$

$w = 6$

$d = 3$

Network Metrics

- Network centralities*
  - Which nodes are more "central" than others?
  - Central nodes are those in the "thick of things" or "focal" among the nodes*.



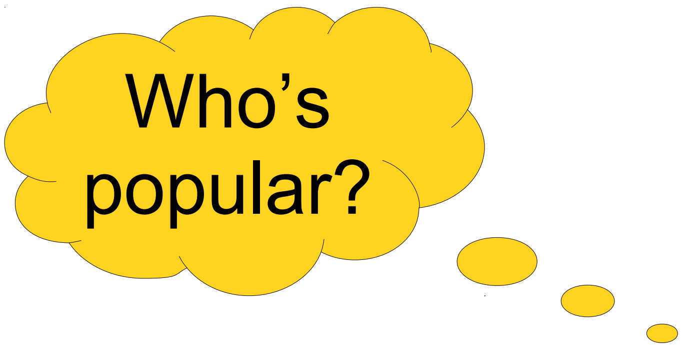
Here, Node A can be considered central because it:

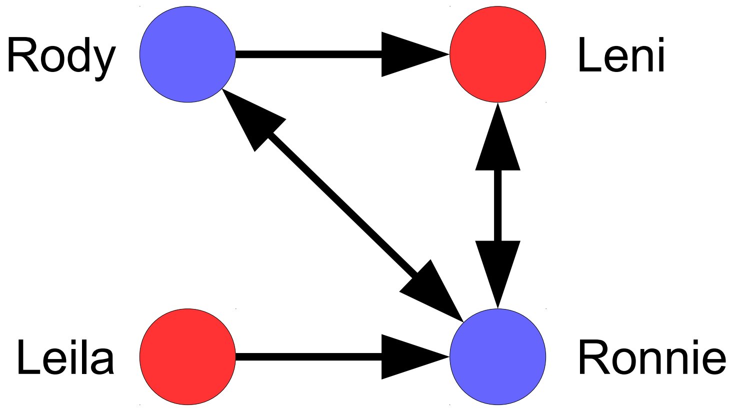
- has more ties;
- can reach others through one edge, while others need two edges;
- can control the flow of data to other nodes.

* *Freeman LC. 1978. Centrality in social networks: Conceptual clarification.* ***Social Networks*** *1, 215-239.*

Network Metrics

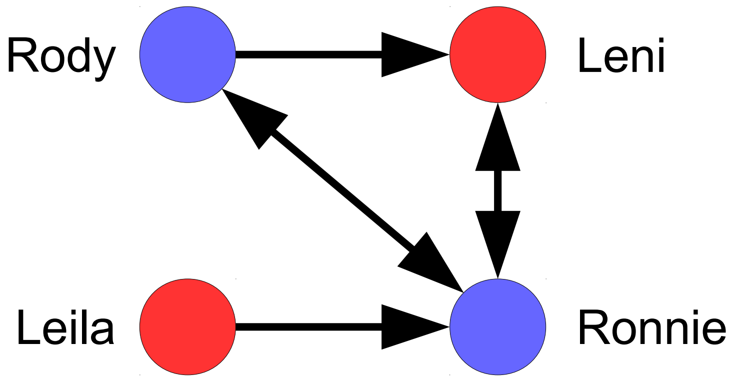
- Network centralities
  - **Degree centrality** – Number of links a node has
    - Concept 1: For undirected network, immediate risk of a node for catching whatever is flowing in the network (gossip, information, virus, etc)
    - Concept 2: For directed network, in-degree (number of edges that point towards the node, i.e., *popularity*) and out-degree (number of edges the point away from the node, i.e., *gregariousness*)





Network Metrics

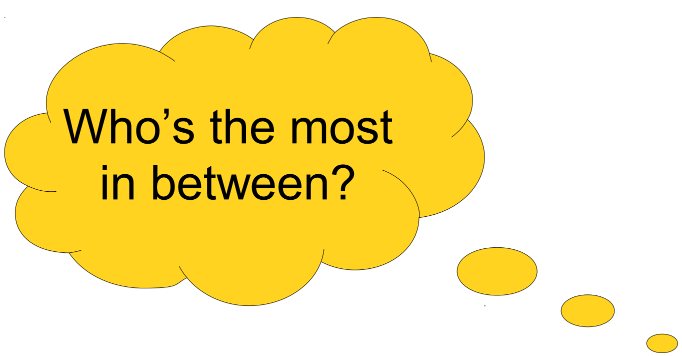
- Network centralities
  - **Closeness centrality*** – Inverse of farness, which is the sum of distances to all other nodes.
    - Computational idea: Compute the shortest distance** between all pairs of nodes
    - Distance is the number of (directed) paths to take to reach another node from a given node.

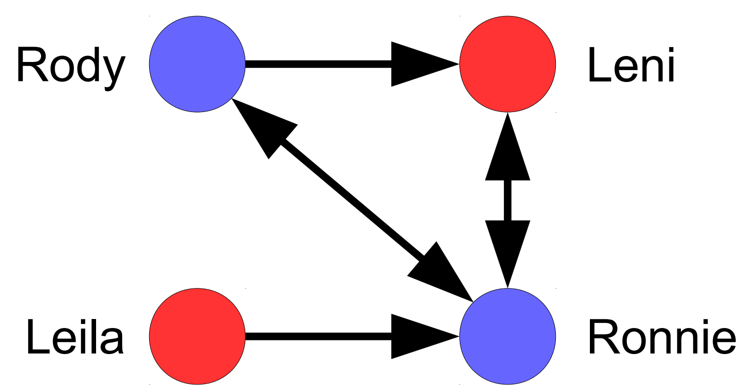


* *Freeman LC. 1978. Centrality in social networks: Conceptual clarification.* ***Social Networks*** *1, 215-239.*
** *Dijkstra EW. 1959. A note on two problems in connexion with graphs.* ***Numerische Mathematik*** *1, 269-271.*

Network Metrics

- Network centralities
  - **Betweenness centrality*** – The extent for which a node is a <u>part of transactions</u> among other nodes.
    - In pinoy's red tape parlance, these are the fixers, go-betweeners, or *tulay*.
    - Intuitive computation is via Dijkstra's algorithm but a faster** one exists.





*Freeman LC. 1978. Centrality in social networks: Conceptual clarification. **Social Networks** 1, 215-239.*

*Brandes U. 2001. A Faster Algorithm for Betweenness Centrality. **Journal of Mathematical Sociology** 25, 163-177.*

Network Metrics

- Characterizing Nodes
  - Hubs and Authorities* (iterative definition)
    - Authorities are nodes that are sources of authoritative information. A good authority is one that is pointed to by many good hubs.
    - Hubs are nodes that are sources of authorities. A good hub is one that points to many good authorities.
    - Can only be performed on a directed network

* *Manning CD, Raghavan P and Schütze H. 2008.* ***Introduction to Information Retrieval****. Cambridge University Press.*

Network Metrics

- Workshop #1
  - Install Pajek into your PC Compatibles (Intel-based chipset running MS-Windows OS)
  - Prepare your data set
    - Using any word processor that can save an ASCII file, format the data file as follows:
    - Line 1: `*Vertices` <number of vertices, *n*>
    - Line 2 to Line *n*+1: <unique integer> "<name of vertex>"

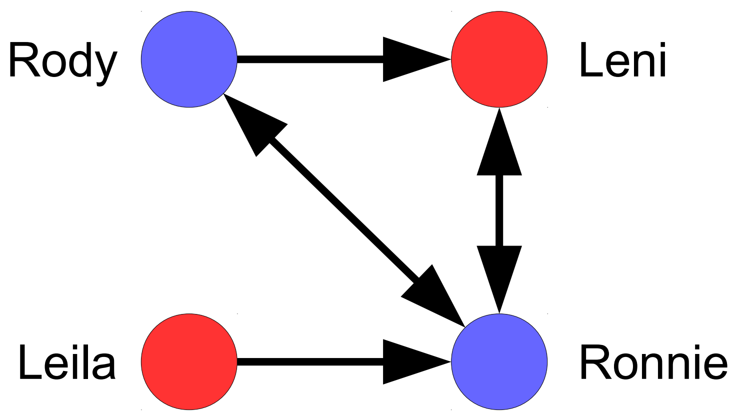
Network Metrics

- Workshop #1
  - Install Pajek into your PC Compatibles (Intel-based chipset running MS-Windows OS)
  - Prepare your data set
    - Using any word processor that can save an ASCII file, format the data file as follows:
    - Line 1: `*Vertices` <number of vertices, *n*>
    - Line 2 to Line *n*+1: <unique integer> "<name of vertex>"
    - Line *n*+2:
      - If Undirected: `*Edges`
      - If Directed: `*Arcs`
    - Line *n*+3 and onward: <integer> <integer>

Network Metrics

- Workshop #1: Example data set: *perception of close friendship*
  - Line 1:
  - Line 2:
  - Line 3:
  - Line 4:
  - Line 5:
  - Line 6:
  - Line 7:
  - Line 8:
  - Line 9:
  - Line 10:
  - Line 11:
  - Line 12:

```
*Vertices 4
1 "Rody"
2 "Leni"
3 "Ronnie"
4 "Leila"
*Arcs
1 2
1 3
2 3
3 1
3 2
4 3
```



*

Network Metrics

- Workshop #1: Example data set: *perception of close friendship*
  - Line 1: `*Vertices 4`
  - Line 2: `1 “Rody”   x_fact 1 y_fact 1 ic Blue`
  - Line 3: `2 “Leni”   x_fact 1 y_fact 1 ic Red`
  - Line 4: `3 “Ronnie” x_fact 1 y_fact 1 ic Blue`
  - Line 5: `4 “Leila”  x_fact 1 y_fact 1 ic Red`
  - Line 6: `*Arcs`
  - Line 7: `1 2`
  - Line 8: `1 3`
  - Line 9: `2 3`
  - Line 10: `3 1`
  - Line 11: `3 2`
  - Line 12: `4 3`

Add these to Lines 2 through 5, respectively

*

Network Metrics

- Workshop #1: Example data set: *perception of close friendship*
  - Draw the Network and explore the drawing options
  - Compute for the following centralities:
    - Degree
    - Closeness
    - Betweenness
  - Identify the Hubs and Authorities

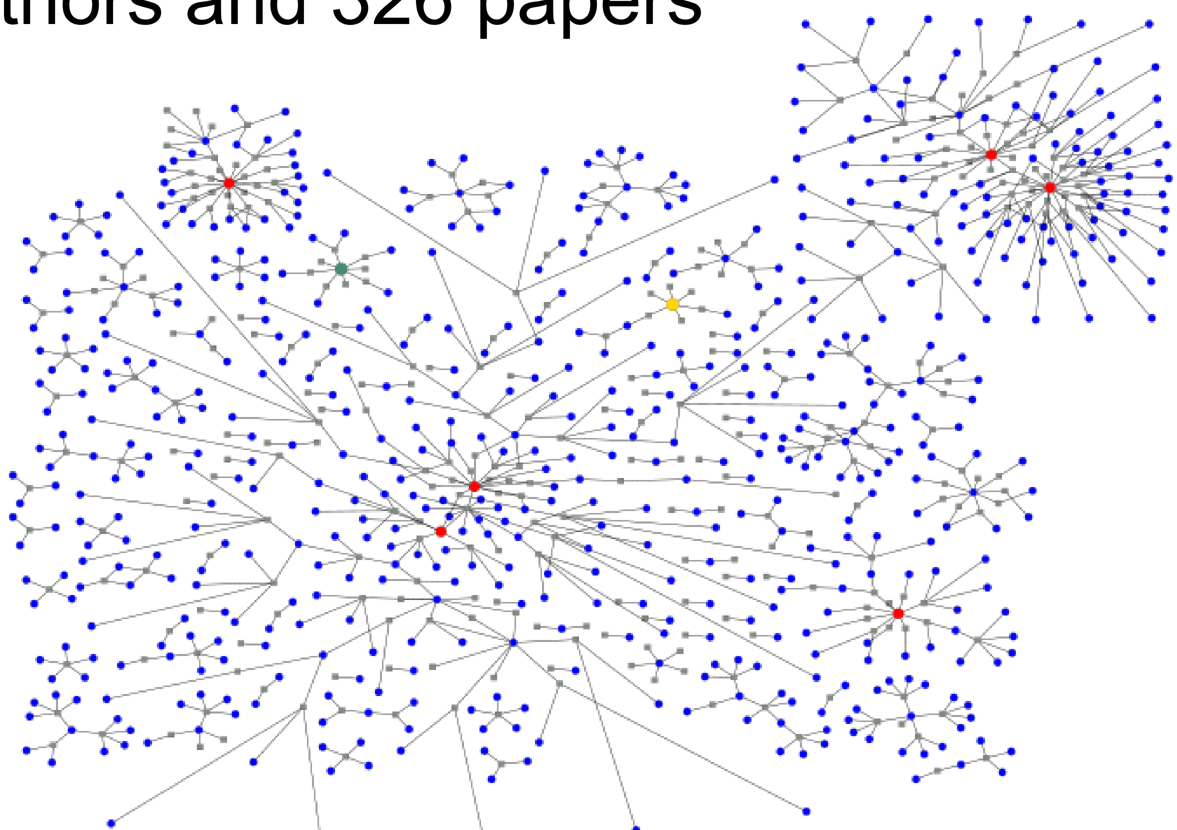
Network Metrics

- Workshop #2: Scientific Collaboration Network
  - Copy the Pajek dataset for the Collaboration Network of Filipino Computer Scientists*
  - Inspect, using your word or text processor (preferably notepad or better), if the data file follows the Pajek input format
    - 542 nodes/authors
    - 969 edges/co-authorship

* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

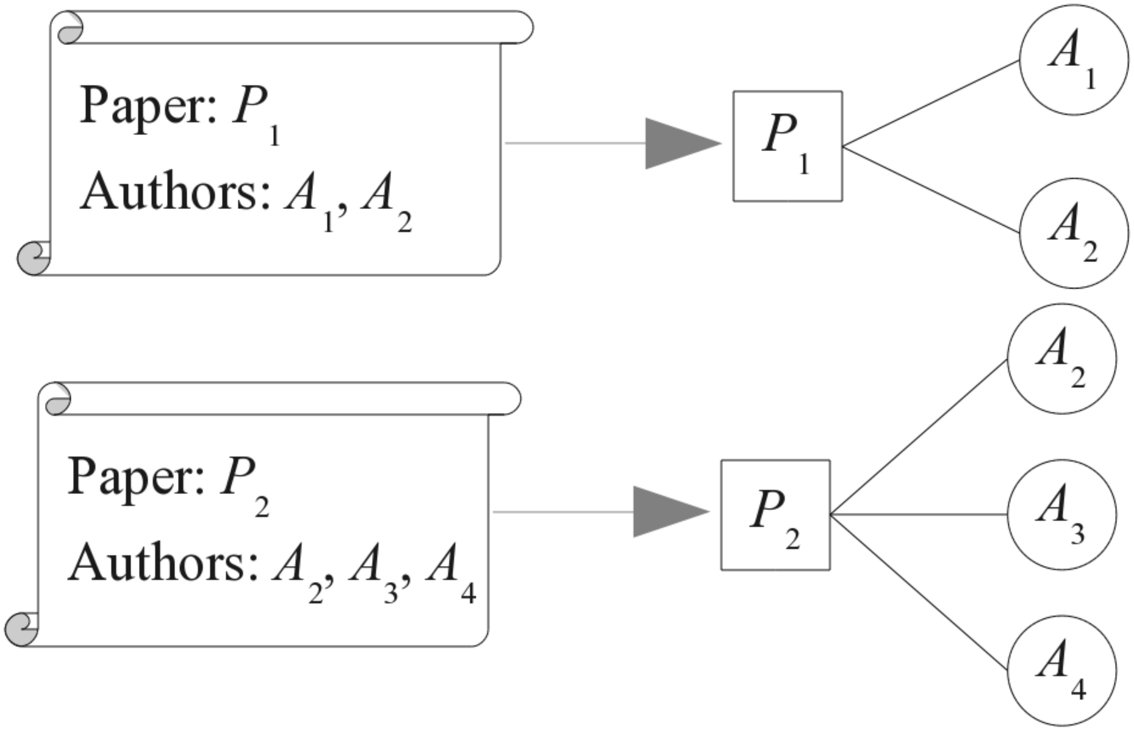
- Workshop #2: Scientific Collaboration Network
  - Started out as a paper-author bipartite network with 542 authors and 326 papers



* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

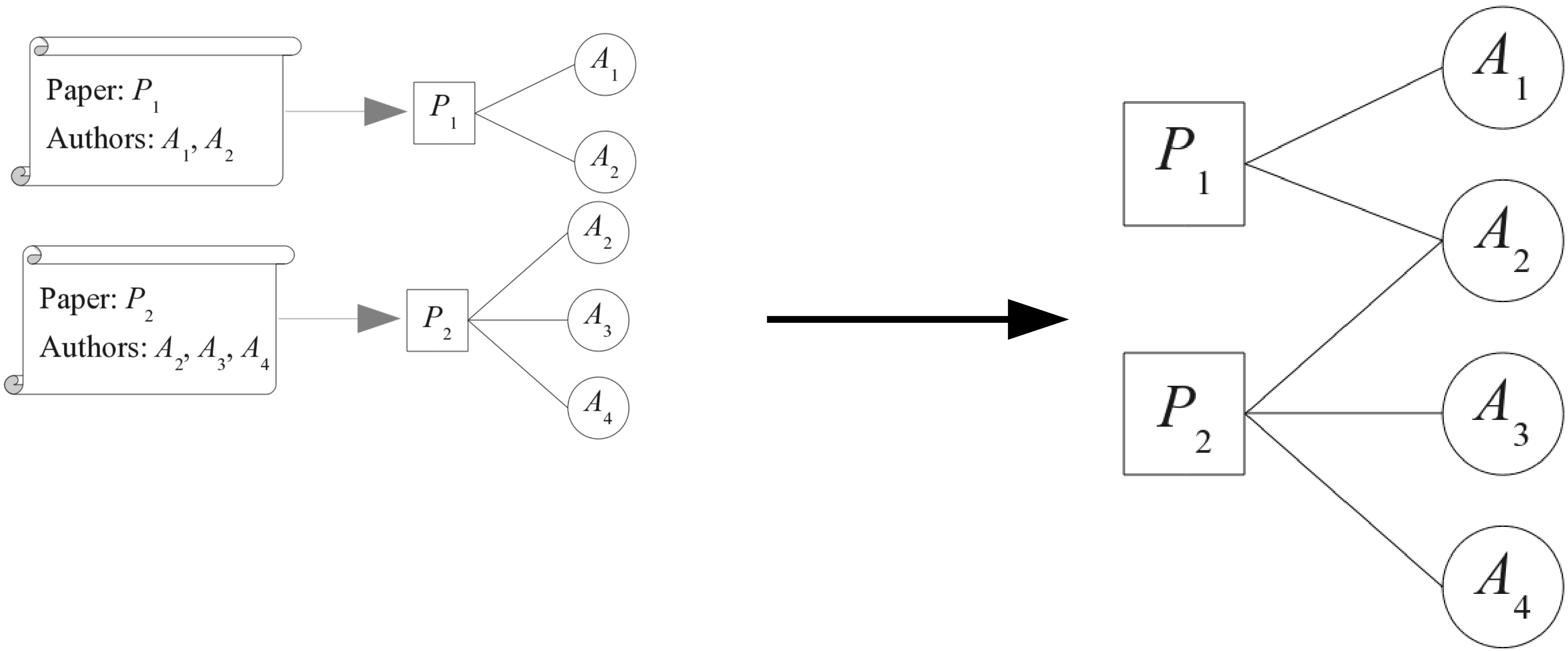
- Workshop #2: Scientific Collaboration Network
  - Idea:



* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

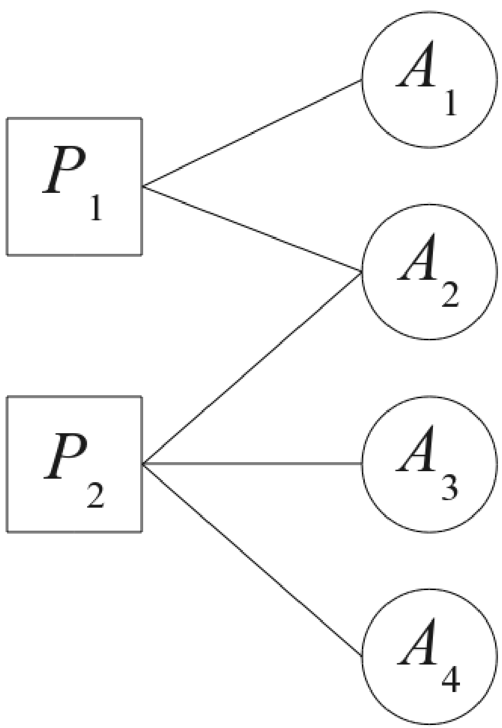
- Workshop #2: Scientific Collaboration Network
  - Idea:



* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

- Workshop #2: Scientific Collaboration Network
  - Idea:



$$
\begin{array}{cc}
 & \begin{array}{cccccc} P_1 & P_2 & A_1 & A_2 & A_3 & A_4 \end{array} \\
\begin{array}{c} P_1 \\ P_2 \\ A_1 \\ A_2 \\ A_3 \\ A_4 \end{array} &
\begin{bmatrix}
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

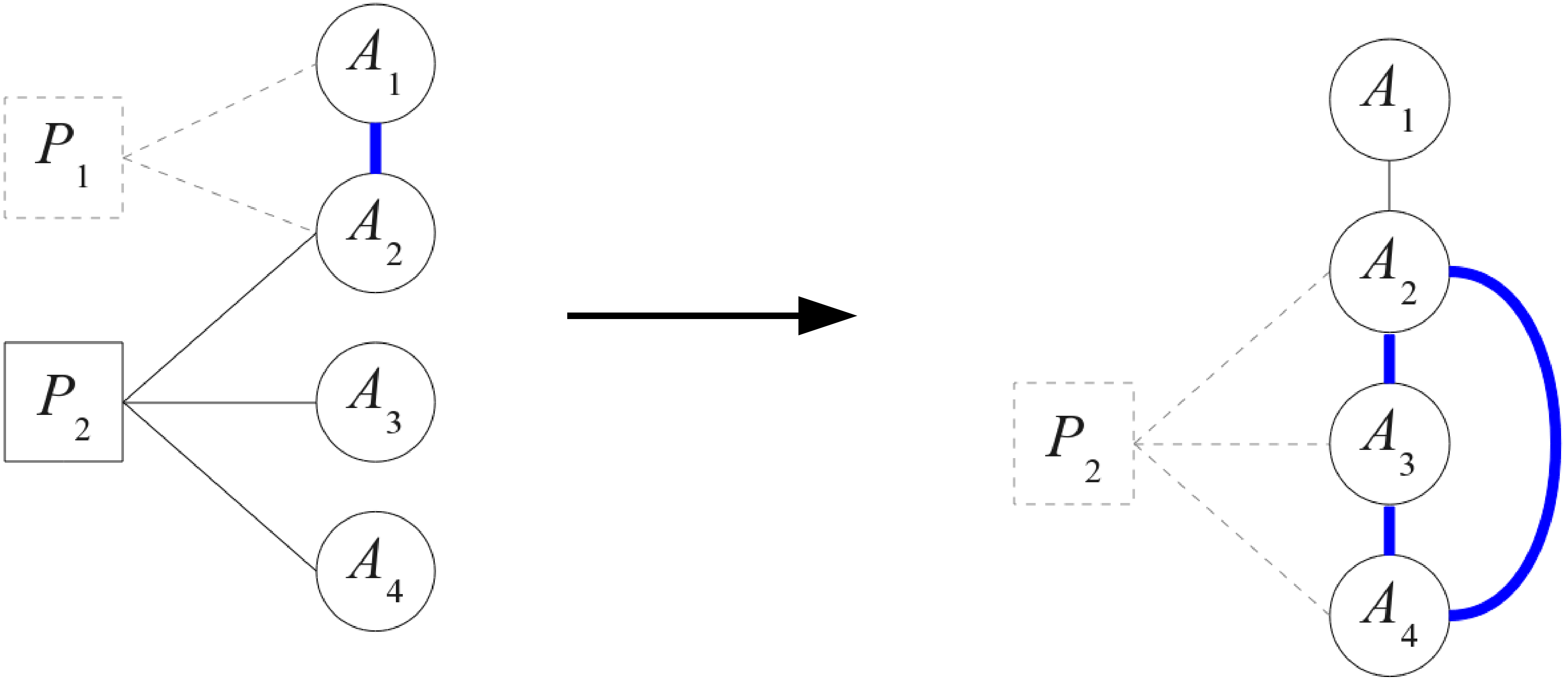
- Workshop #2: Scientific Collaboration Network
  - Idea:

$$\begin{array}{c|cccccc|} & P_1 & P_2 & A_1 & A_2 & A_3 & A_4 \\ \hline P_1 & 0 & 0 & 1 & 1 & 0 & 0 \\ P_2 & 0 & 0 & 0 & 1 & 1 & 1 \\ A_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ A_2 & 1 & 1 & 0 & 0 & 0 & 0 \\ A_3 & 0 & 1 & 0 & 0 & 0 & 0 \\ A_4 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline \end{array} \longrightarrow \begin{array}{c|cccc|} & A_1 & A_2 & A_3 & A_4 \\ \hline P_1 & 1 & 1 & 0 & 0 \\ P_2 & 0 & 1 & 1 & 1 \\ \hline \end{array}$$

* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

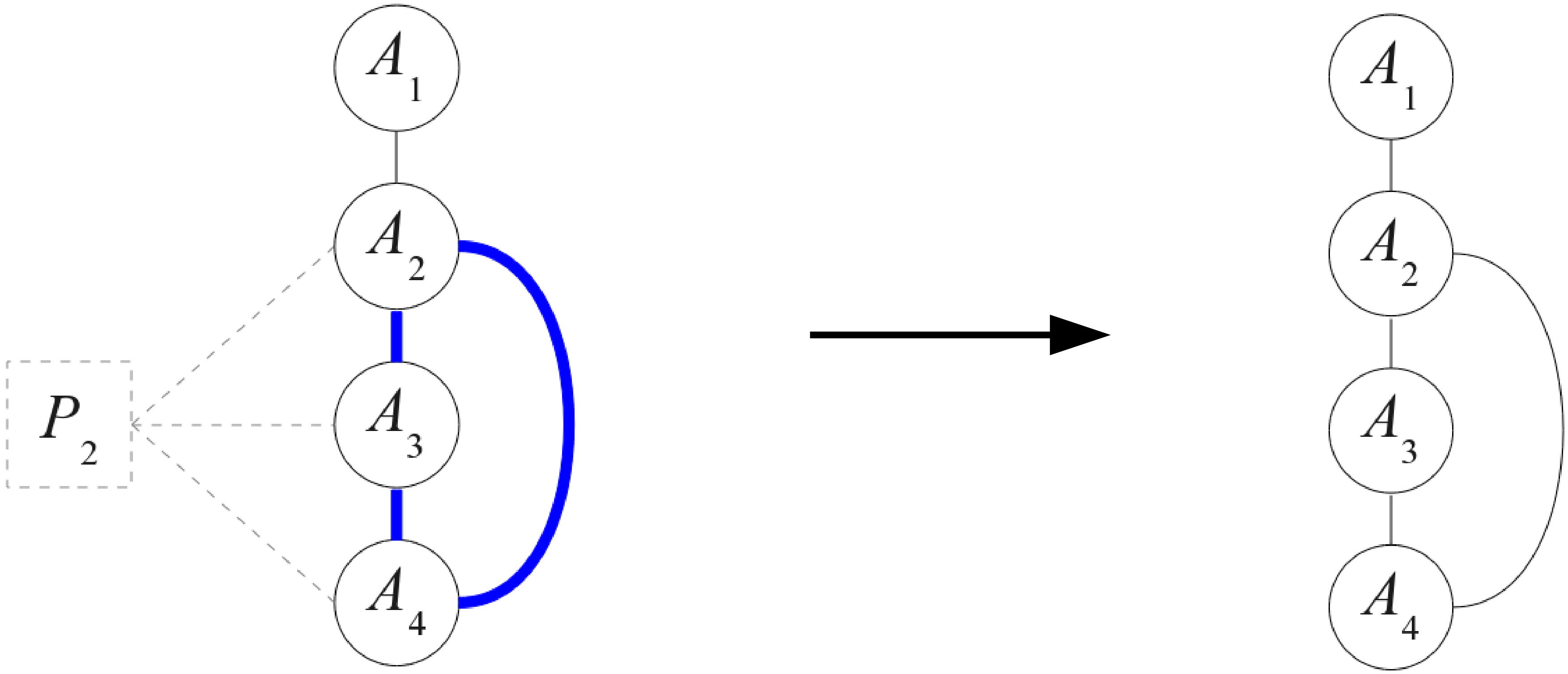
- Workshop #2: Scientific Collaboration Network
  - Idea:



* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

- Workshop #2: Scientific Collaboration Network
  - Idea:



* *Pabico JP. 2010. Authorship patterns in computer science research in the Philippines.* ***Philippine Computing Journal*** *5(1):1-13*

Network Metrics

- Workshop #2: Scientific Collaboration Network
  - Load the dataset
  - Draw the network using various drawing techniques
  - Compute for the centralities
    - Degree
    - Closeness
    - Betweenness
  - Characterize the nodes
    - Hubs and authorities

Network Metrics

- Workshop #3: Sex Network*
  - Load the dataset Sex Network
    - Vertices: 16,730
    - Edges: 50,632
  - The data is composed of two types of nodes:
    - Male who are escort service seeker; and
    - Female who are escort service provider.
  - Modify the data so that it can be accepted by Pajek
    - Can we just use your word/text processor?

* *Rocha LEC, Liljeros F and Holme P. 2010. Information dynamics shape the sexual networks of Internet-mediated prostitution.* ***Proceedings of the National Academy of Sciences of USA*** *107(13):5706--5711.*

Network Metrics

- Workshop #3: Sex Network
  - Draw the network using various drawing techniques
  - Compute for the centralities
    - Degree
    - Closeness
    - Betweenness
  - Can we find who are the hubs and the authorities?

Network Metrics

- Workshop #4: Doctors Network*
  - 241 vertices, 1098 edges
  - Draw the network using various drawing techniques
  - Compute for the centralities
    - Degree
    - Closeness
    - Betweenness
  - Can we find who are the hubs and the authorities?

* *Coleman J, Katz E and Menzel H. 1957. The diffusion of an innovation among physicians.* ***Sociometry*** *20(4):253-270.*

Questions?

- Email to <jppabico@uplb.edu.ph> for:
  - Questions requiring detailed answers
  - Proposals for research collaboration
    - Soft computing and machine learning
    - HPC/scheduling and dynamic load balancing
    - Wireless adhoc networks
    - Computer security and forensics
    - Information visualization
- http://www.ics.uplb.edu.ph/jppabico