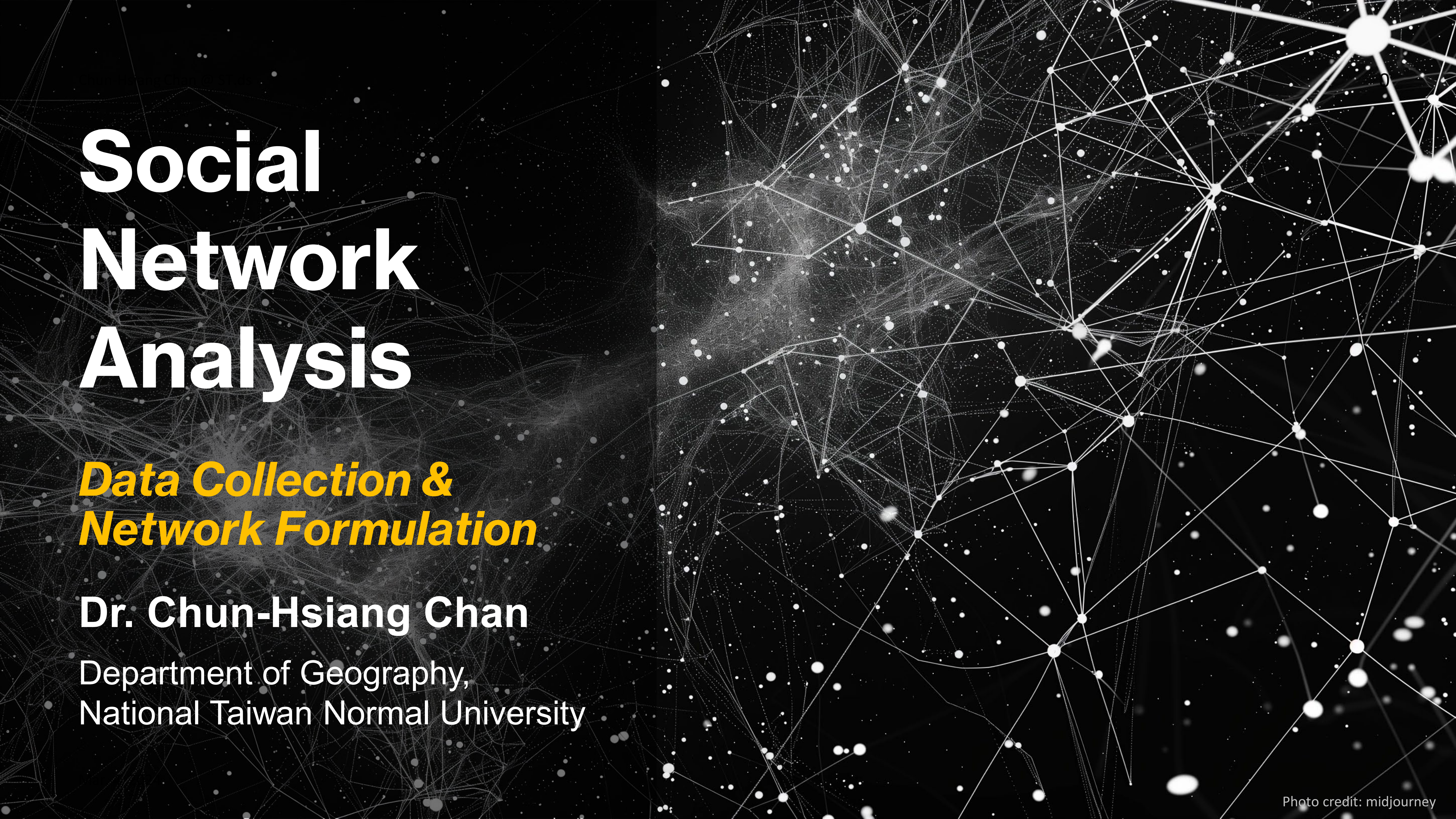


Social Network Analysis

Data Collection & Network Formulation

Dr. Chun-Hsiang Chan

Department of Geography,
National Taiwan Normal University

Photo credit: midjourney

Outline

- Network Data
- Boundary Specification and Sampling
- Types of Networks
- Network Data Collection
- Network Formulation
- Network in Matrix
- Network Characteristics
- Paper Reading
- References

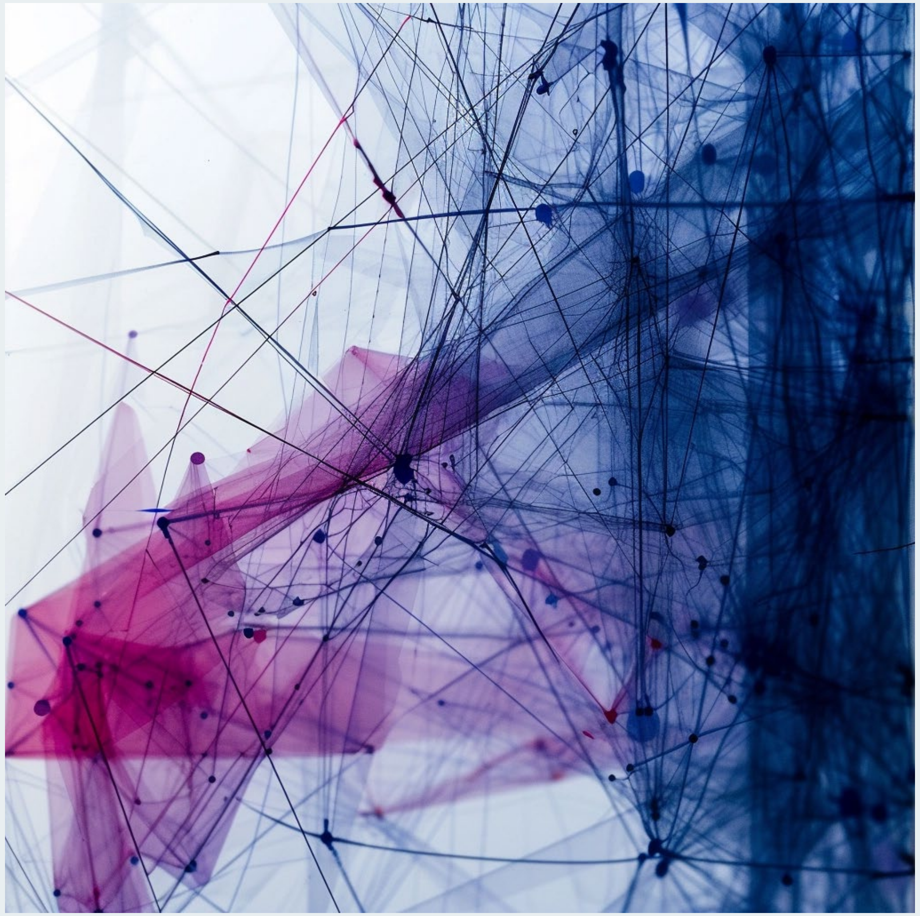


Photo credit: midjourney

Network Data

- Before collecting the network data, we need to understand the formulation of network data. Wasserman and Faust (1994) mentioned that social network data consists of at least one structural variable measured on a set of actors.

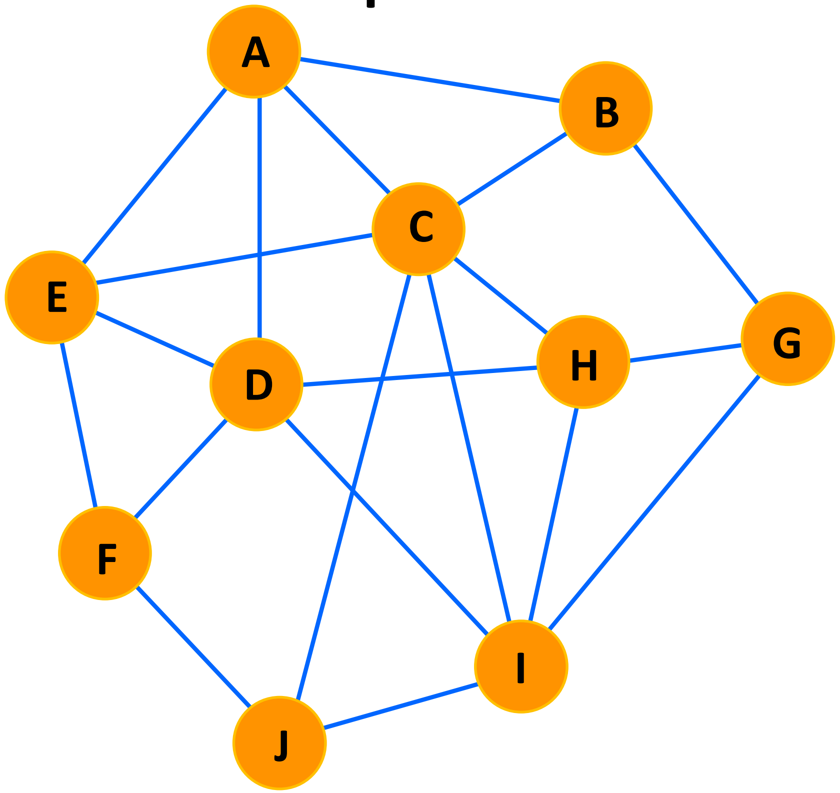
Network Data

– **Structural variables** are measured on pairs of actors and are the cornerstone of the social network data set. Structural variables measure ties of a specific kind between pairs of actors.

– **Composition variables** are measurements of actor attributes. Composition variables, or actor attribute variables, are of the standard social and behavioral science variety and are defined at the level of individual actors.

Network Data

– **Structural and composition variables**

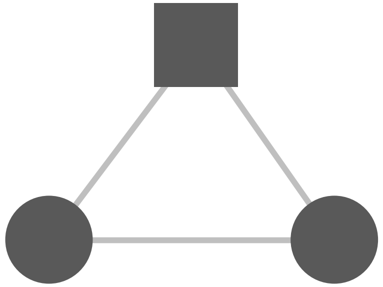


Actor (Node)	Name	Grade
A	Jimmy	1
B	Lilly	1
C	Helen	2
D	Gloria	4
E	Wendy	3
F	Cindy	1
G	Joy	2
H	Nick	3
I	Tiffany	3
J	Mike	4

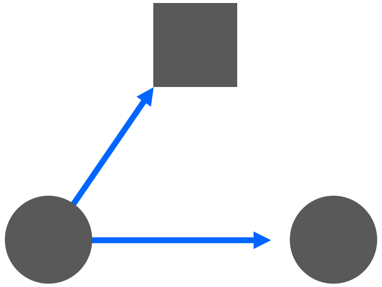
Network Data

Modes

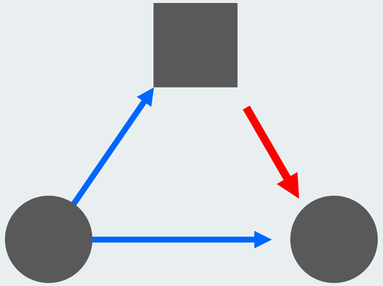
- The typical modes in the network analysis are one-mode and two-mode networks.



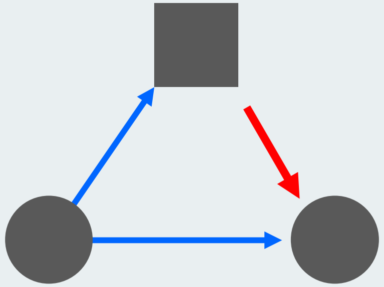
One-mode tie



One-mode tie



One-mode tie →
Two-mode tie



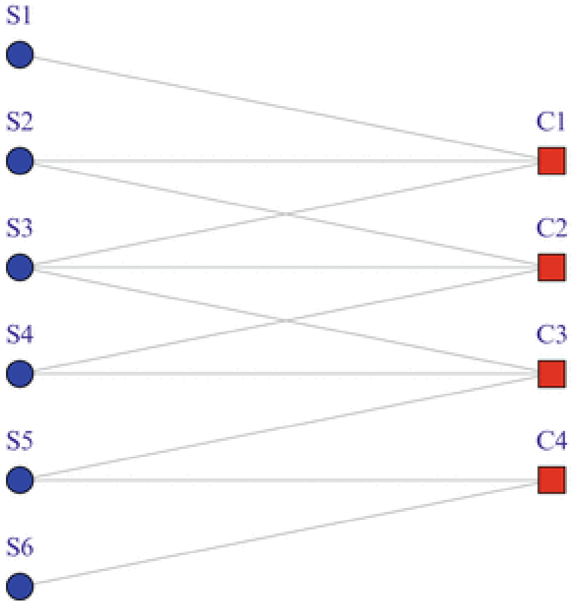
Two-mode tie →
One-mode tie

Network Data

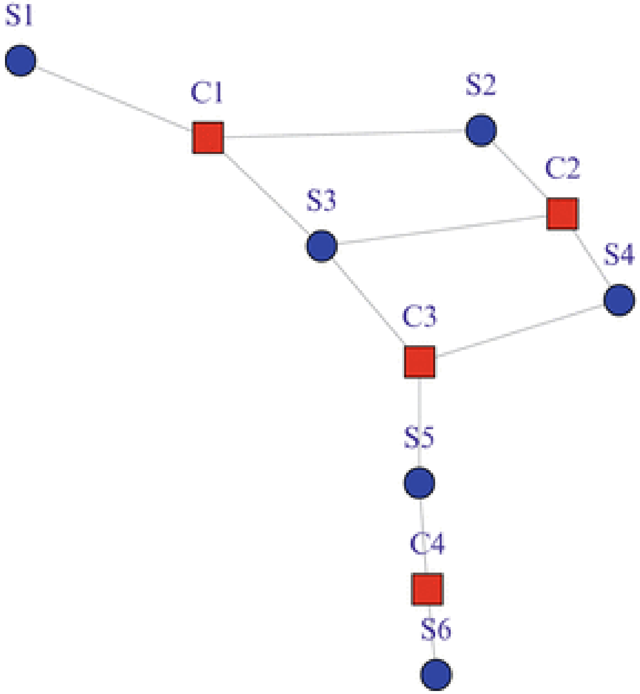
Affiliation variables

- An affiliation network is a specific type of two-mode network that arises in social network studies. Affiliation networks are two modes but have only one set of actors. The second mode in an affiliation network is a set of events (such as clubs or voluntary organizations) to which the actors belong.

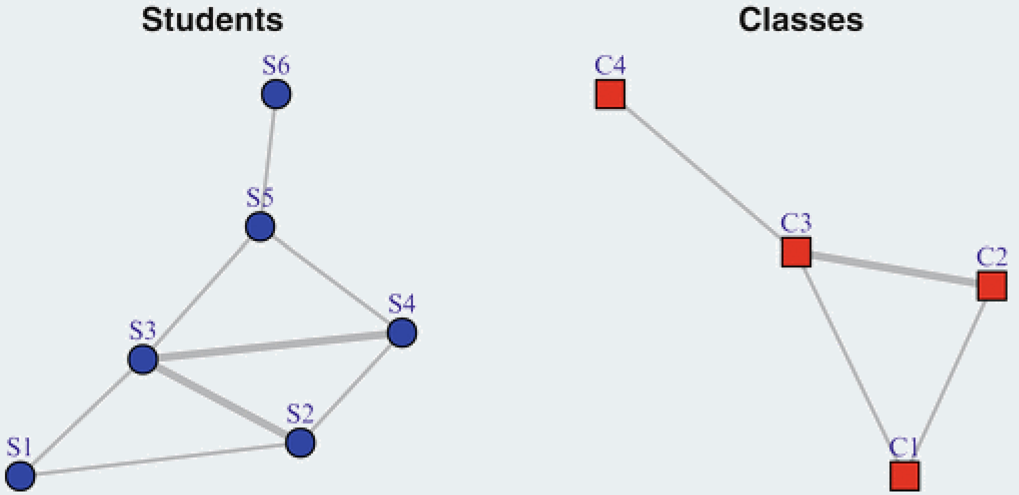
Network Data



Affiliation network as bipartite graph



Simple plot of affiliation network



Plots of one-mode projections

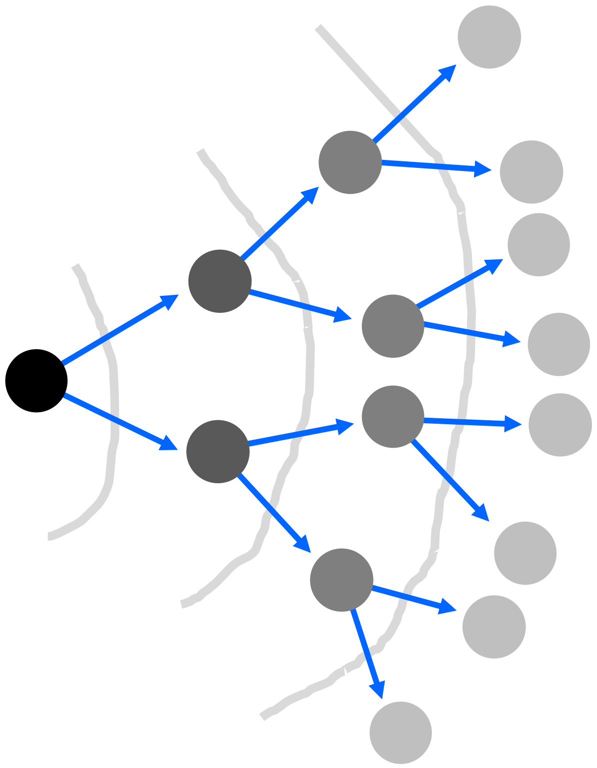
Luke, D., & Luke, D. A. (2015). Affiliation networks. *A User's Guide to Network Analysis in R*, 125-144.

Boundary Specification and Sampling

Population

– At the beginning of data collection, the first question is how to define the boundary conditions of network data. If your population is small, you may collect the entire data. But if your population is large, then how do you define your boundary?

– **Think about it!**



Boundary Specification and Sampling

Snowball sampling

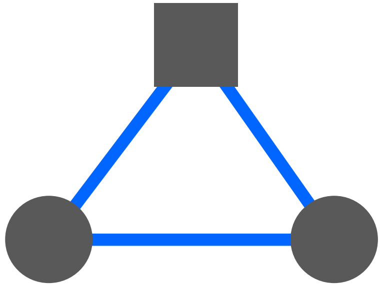
Mar. 4, 2024

Boundary Specification and Sampling

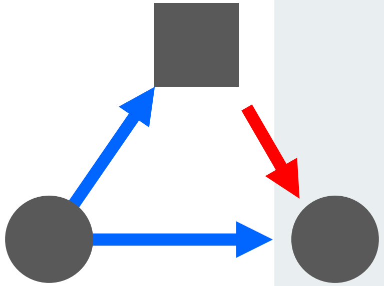
- **Snowball sampling** (or chain sampling, chain-referral sampling, referral sampling)
- **Snowball sampling** is a nonprobability sampling technique where existing study subjects recruit future subjects from among their acquaintances in sociology and statistics research.
- Thus, the sample group is said to grow like a rolling snowball. As the sample builds up, enough data are gathered to be useful for research. This sampling technique is often used in hidden populations, such as drug users or sex workers, which are difficult for researchers to access.

Types of Networks

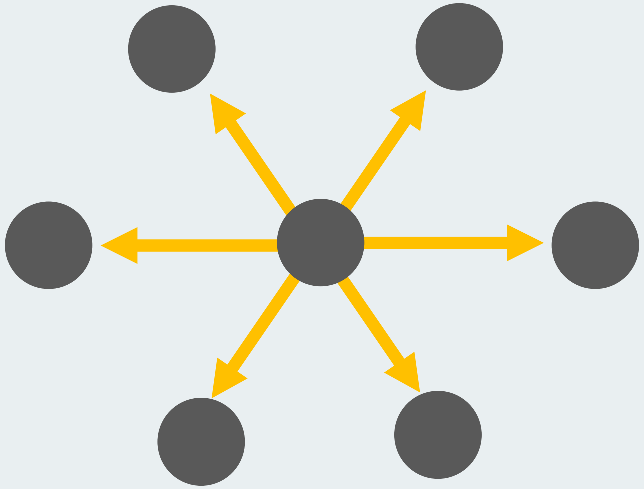
- Various types of networks were conducted in previous studies. The simplest way to categorize networks is **mode**: one-mode and two-mode networks. Some studies even adopted a higher mode to describe the network structure, such as three-mode networks.



One-mode network

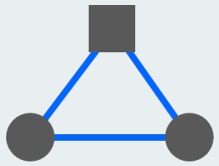


Two-mode network

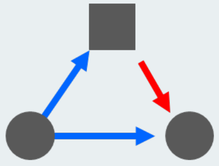


Ego-centered Network

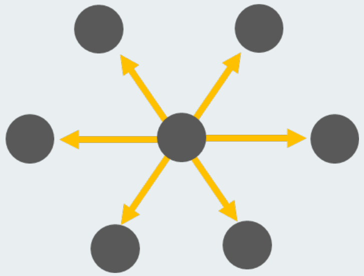
Types of Networks



One-mode network



Two-mode network



Ego-centered Network

Name	One-mode	Two-mode	Ego-centered
Actors	Variety of types, Subgroup, Organization, Company	Two sets of actors who can be of the general types as described for one-mode networks	Ego-centered network consists of a focal actor, termed ego, as set of alters who have ties to ego
Relations	Friendship (Social media) Transaction Interaction Movement Marriage	At least one relation is measured between actors in the two sets.	Measurements on the ties among these alters.

Ego-centered Network

Network Data Collection

– Measurement

Social network data differs from standard social and behavioral science data in a number of important ways. The presence of relations has limitations for a number of measurement issues, including the unit of observation (actor, pair of actors, relational tie, or event), the modeling unit (the actor, dyad, triad, subset of actors, or network), and the quantification of the relations (directed vs undirected; dichotomous vs valued).

Network Data Collection

– **Unit of observation**

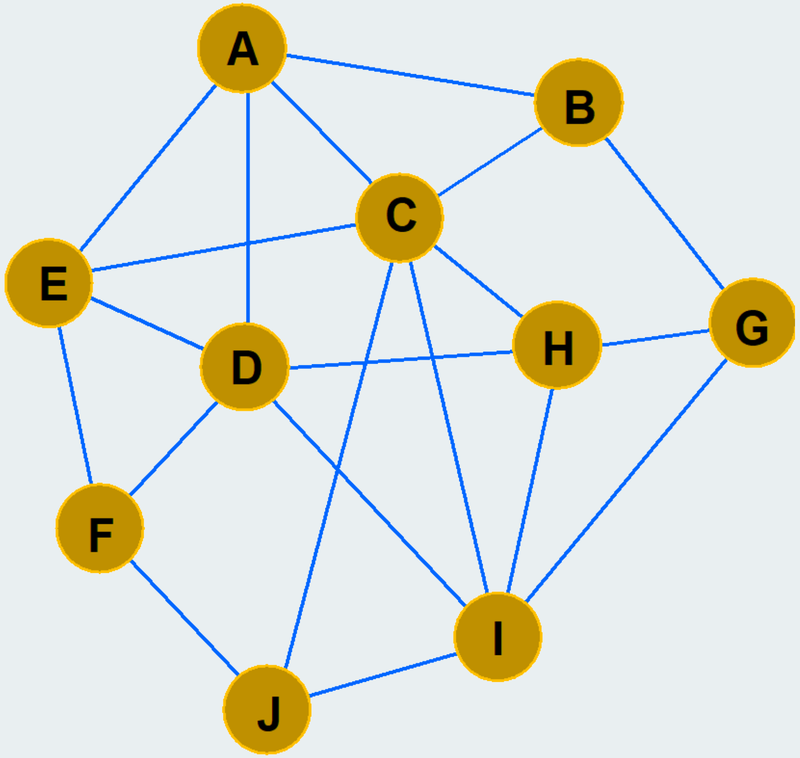
The unit of observation is the entity on which measurements are taken; most often, social network data are collected by observing, interviewing, or questioning individual actors about the ties of these actors to other actors in the set.

– **Modeling Unit**

The modeling unit could be various levels, such as actor, dyad, triad, subgroup, or set of actors or networks.

Network Data Collection

- There are several ways to collect and construct the network data from …
    - Questionnaires
    - Interviews
    - Observations
    - Experiments
    - Archival records
    - Other sources



Zhou, D., Yan, Z., Fu, Y., & Yao, Z. (2018). A survey on network data collection. *Journal of Network and Computer Applications*, *116*, 9-23.

Network Data Collection

Questionnaires

Alter traits obtained include:

- sex
- age (in categories)
- race/ethnicity – 5 categories, allows for multiple choices
- religious affiliation
- educational attainment
- alter’s relation to ND (e.g. student, roommate, same dorm)
- support received (financial, comfort, social, advice)
- assessments by the respondent (ego) of the alter’s
  - health (waves 3, 4, and 5)
  - happiness (waves 3, 4, and 5)
  - physical activity (waves 4-8)
  - sleep (waves 3, 4, and 5)

Relationship (tie) traits:

- Relationship type
  - Family (parent, step parent, sibling, step sibling, other family)
  - Romantic partner
  - Friend
- Closeness
- Trust
- Relationship duration (in years)
- Frequency of interaction (daily, weekly, monthly, less often)
- Subjective similarity assessment (waves 1 & 2 only)
- Similarity to ego on 5 activities (specified by ego)

Network Data Collection

– **Interviews**

General name generator questions were used to compile a “network partner” list.

This list was then narrowed down to identity two categories of people:

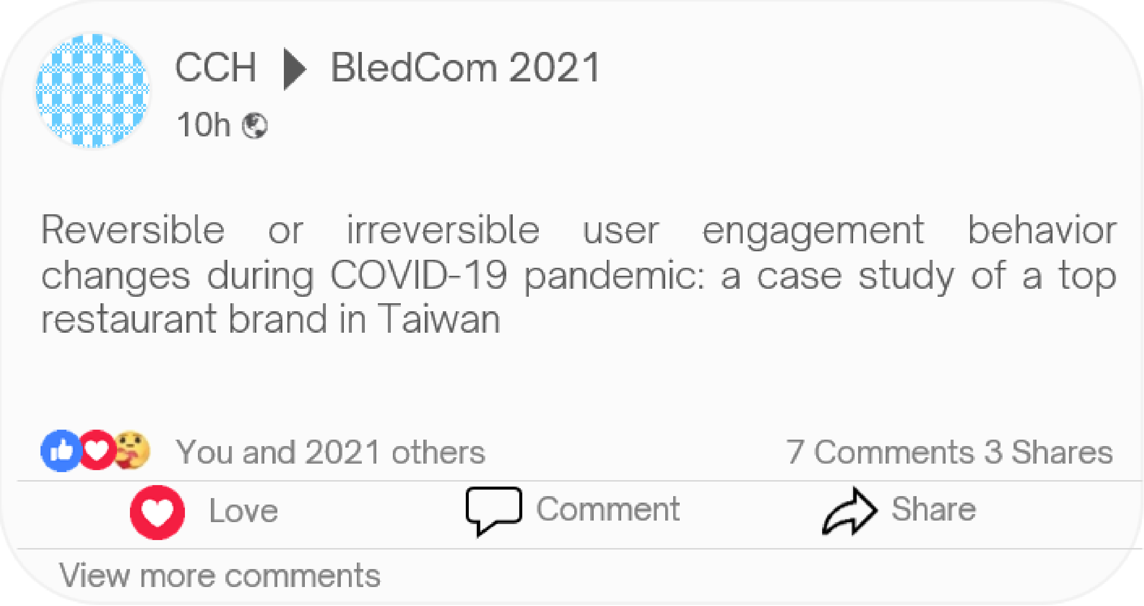
1. Decisionmakers regarding the daughter, and
2. Core influencers (up to 5 people with most influence).

Details of the interaction were used to determine whether the information involved the following types of social support:

1. Exchange of information or advice
2. Providing direct care
3. Offering or requesting material support
4. Offering emotional support.

Network Data Collection

– Observations – Social Media



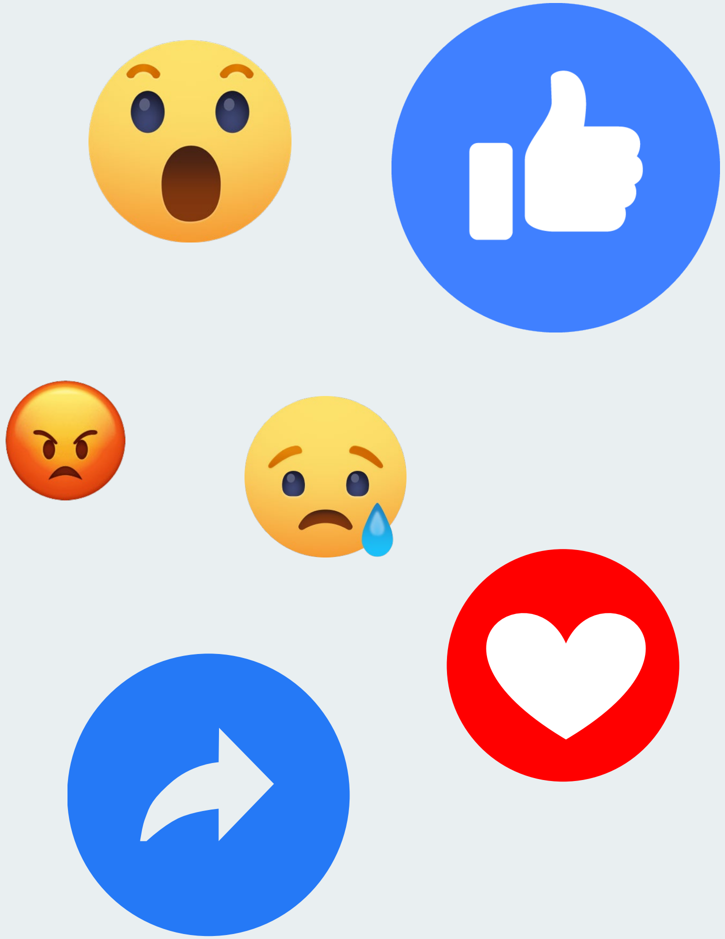
CCH ▶ BledCom 2021

10h

Reversible or irreversible user engagement behavior changes during COVID-19 pandemic: a case study of a top restaurant brand in Taiwan

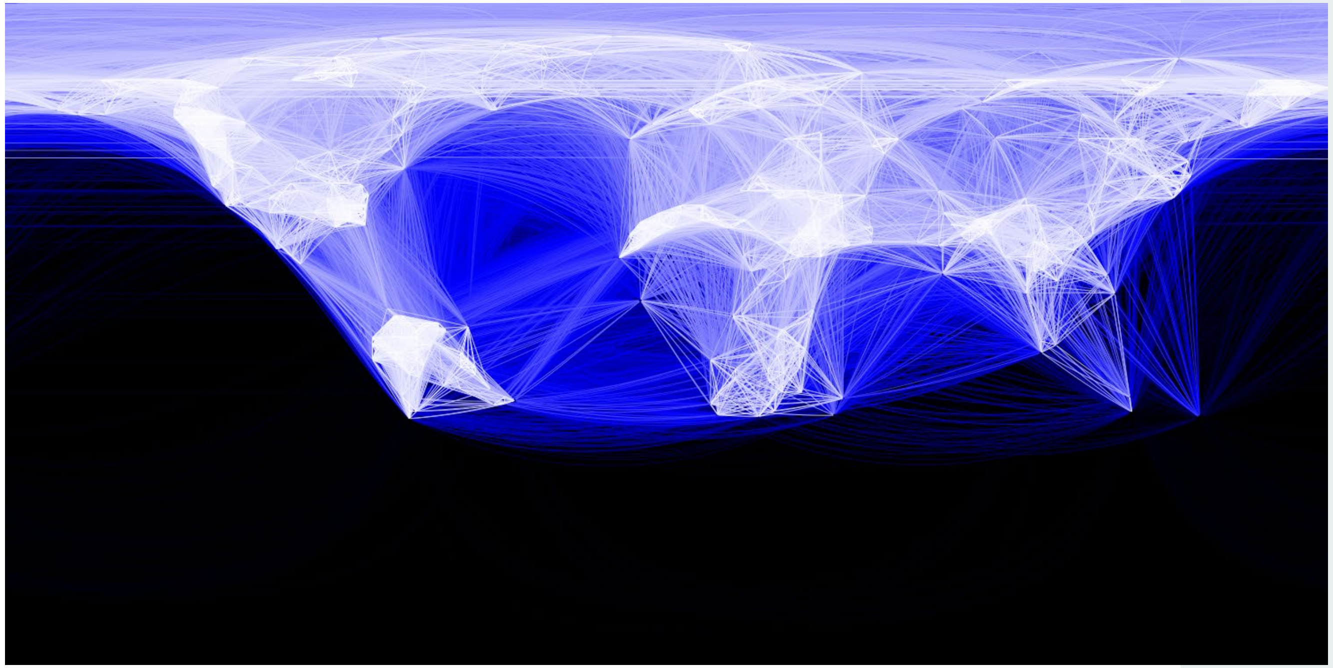
You and 2021 others

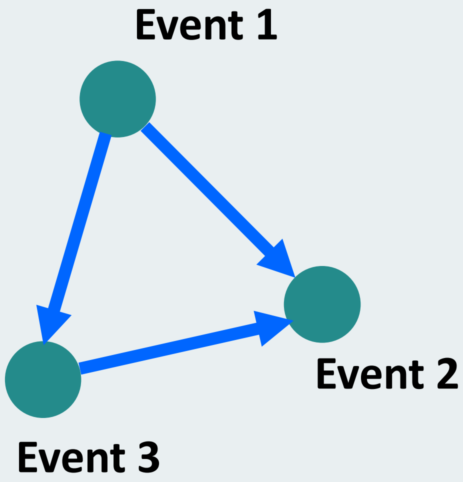
7 Comments 3 Shares

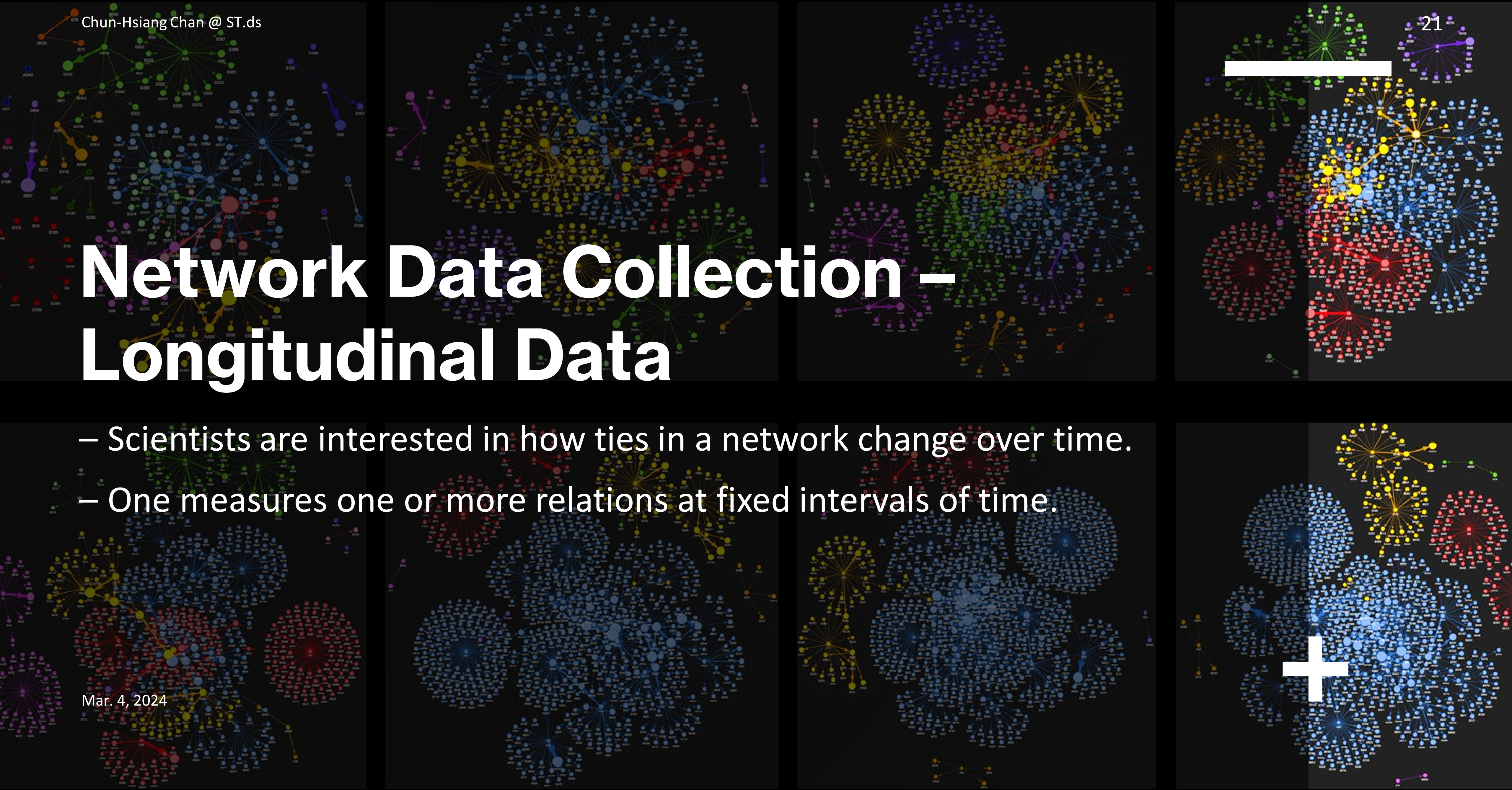


Network Data Collection

– Experiments – Heatwave Network







Network Data Collection – Longitudinal Data

- Scientists are interested in how ties in a network change over time.
- One measures one or more relations at fixed intervals of time.

Measurement – Validity, Reliability, Accuracy, Error

- **Validity:** a measure of a concept is valid to the extent that it measures what it is intended to measure.
- **Reliability:** a measure of a variable or concept is reliable if repeated measurements give the exact estimates of the variable.
- **Accuracy:** sociometric data are collected by having people report on their interactions with other people.
- **Error:** Measurement error occurs when there is a discrepancy between the true score or value of a concept and the observed (measured) value of that concept.

Network Formulation

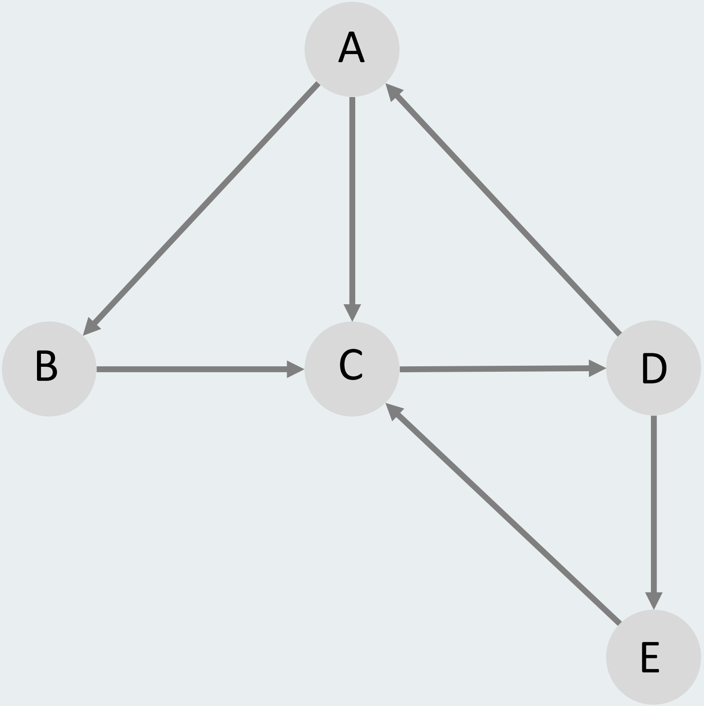
- One way of conceptualizing networks mathematically is as graphs. The term ‘graph’ here does not refer to a diagram but rather a mathematical object (Harary 1969).
- A graph $G(V, E)$ consists of a set of vertices V (also called nodes or points), and a set of edges E (or links or lines).
- The edges connect pairs of vertices. To express that an edge connecting vertices u and v exists in a graph G, we write $(u, v) \in E(G)$.

Network Formulation

- The graph could be **directed** or **undirected** depending on the characteristics of nodal relations.
- Given a **directed** network structure shown as the right-hand-side figure,

$$G(V,E)\ ,where$$
$$V\ =\ \{A,B,C,D,E\}$$
$$E\ =\ \{(A,B),(A,C),(B,C),(C,D),(D,E),(E,C),(D,A)\}$$

- The nodal relation can be ‘is the parent of’ and ‘gives advice to’.

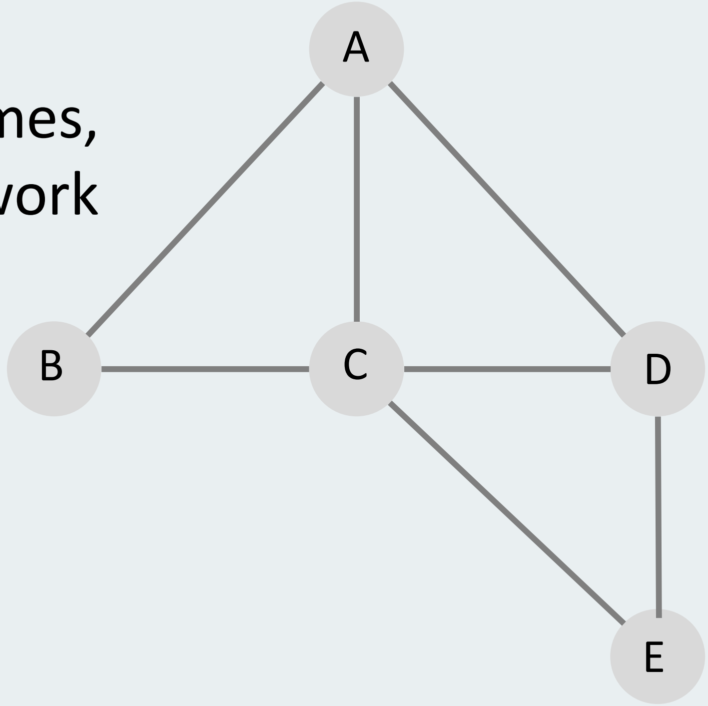


Network Formulation

- Previous example is an **undirected** network, but sometimes, the nodes have no reciprocated relations. Given a network structure shown as right-hand-side figure,

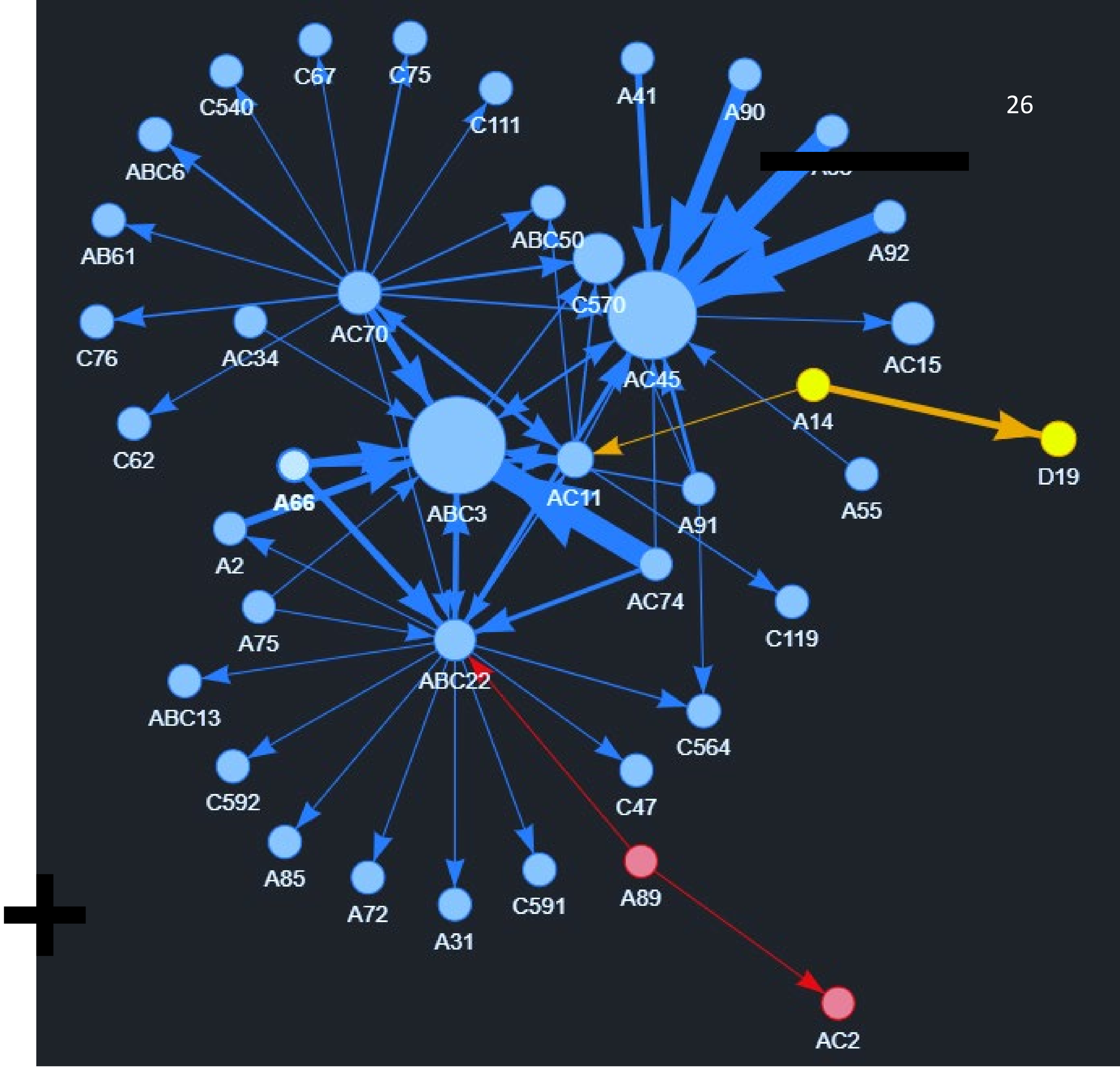
$$G(V, E)\ , where$$
$$V\ =\ \{A, B, C, D, E\}$$
$$E\ =\ \{(A, B), (A, C), (A, D), (B, C), (C, D), (C, E), (D, E)\}$$

- The nodal relation can be ‘was seen with’ or ‘is kin to’.



Network Formulation

- In addition to the directed and undirected networks, we can add weighting parameters for each edge, representing the frequency or strength of nodal relations between two connected nodes.
- As a result, we can divide networks into weighted and unweighted networks.



Network in Matrix

We may use the adjacent matrix to demonstrate the relationship between nodes in the network.

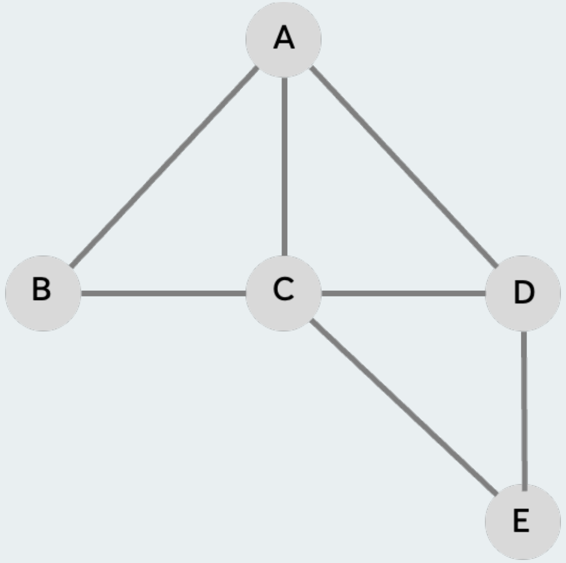
$$G = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \mathit{is\ a\ symmetric\ matrix}$$

If A and B exist a linkage (connection), then the values of [0,1] and [1,0] are 1.
If A and B do not exist a linkage, then the values of [0,1] and [1,0] are 0.

Target

Source

	A	**B**	**C**	**D**	**E**
A	0	1	1	1	0
B	1	0	1	0	0
C	1	1	0	1	1
D	1	0	1	0	1
E	0	0	1	1	0

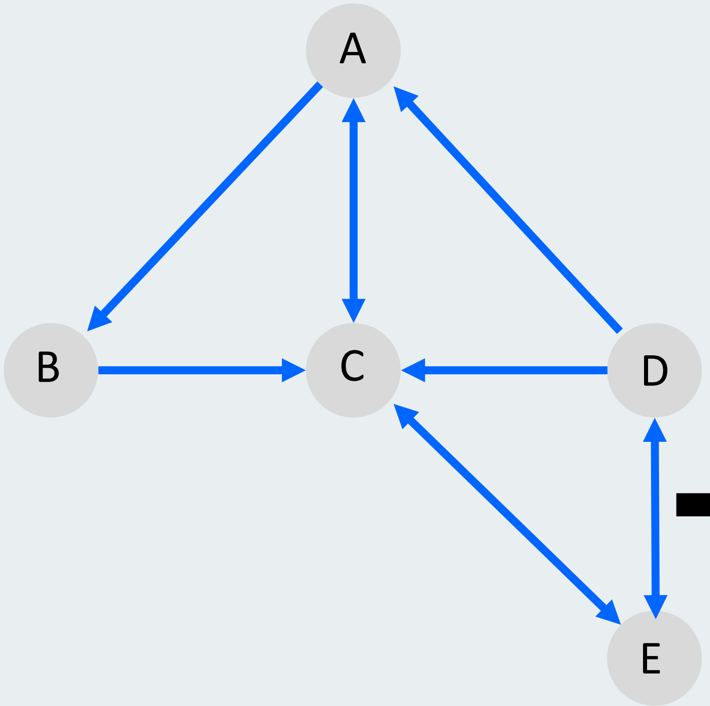


Network in Matrix

$$G = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} is\ an\ asymmetric\ matrix$$

If there exists a linkage (connection) from A to B, then the value of [0,1] is 1; otherwise, the value is 0.

		Target				
		A	**B**	**C**	**D**	**E**
Source	**A**	0	1	1	0	0
	B	0	0	1	0	0
	C	1	0	0	0	1
	D	1	0	1	0	1
	E	0	0	1	1	0



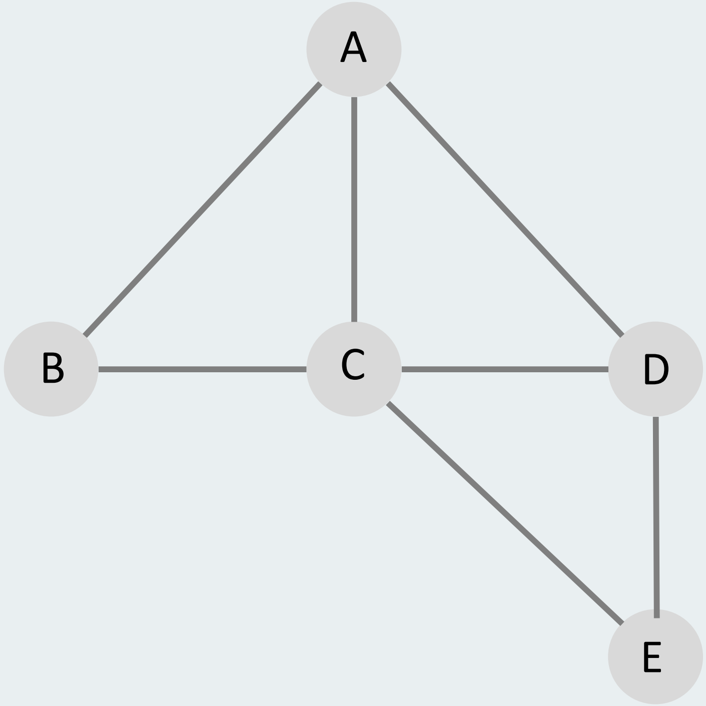
Network in Matrix

– Try it by yourself!

Target

Source

	A	**B**	**C**	**D**	**E**
A					
B					
C					
D					
E					



Network in Matrix

- **Male**; **Female**

Target

Source

	A	B	C	D	E
A	0	1	1	0	0
B	0	0	1	0	0
C	1	0	0	0	1
D	1	0	1	0	1
E	0	0	1	1	0

Block Density Matrix

	M	F
M	0.67	0.17
F	0.50	1.00

Image Matrix
Cut-off (>0.5)

	M	F
M	1	0
F	0	1

Network in Matrix

– Adjacent Matrix Computation

G

Target

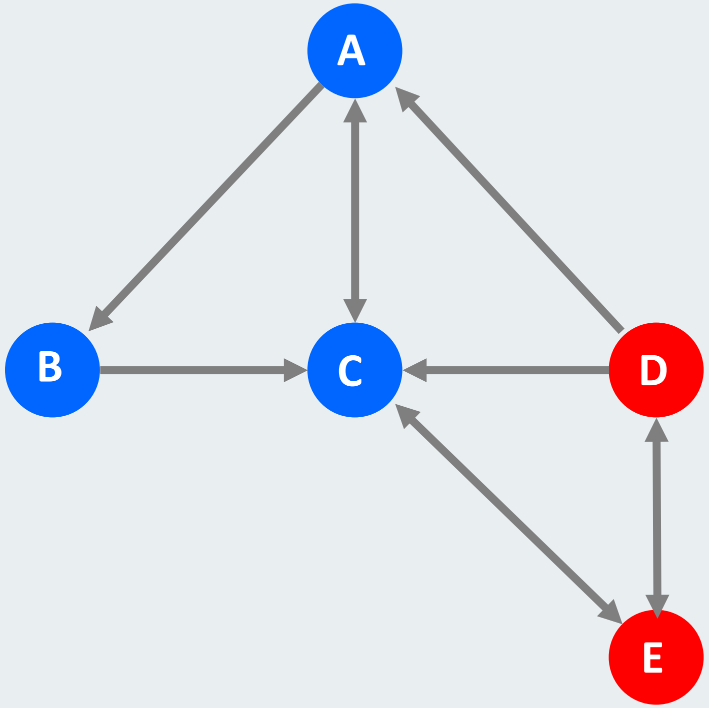
Source

	A	B	C	D	E
A	0	1	1	0	0
B	0	0	1	0	0
C	1	0	0	0	1
D	1	0	1	0	1
E	0	0	1	1	0

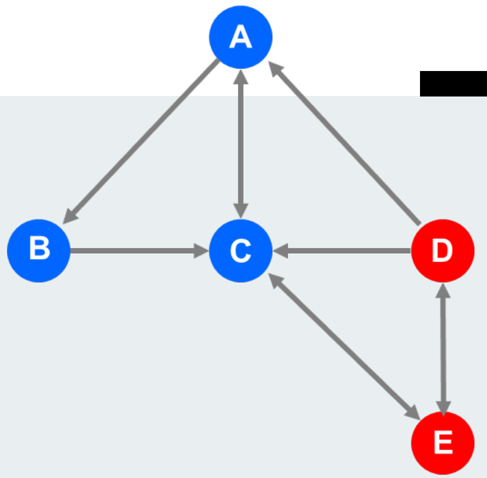
S

Sex

	M	F
A	1	0
B	1	0
C	1	0
D	0	1
E	0	1



Network in Matrix



G

Target

Source

	A	B	C	D	E
A	0	1	1	0	0
B	0	0	1	0	0
C	1	0	0	0	1
D	1	0	1	0	1
E	0	0	1	1	0

×

S

Sex

	M	F
A	1	0
B	1	0
C	1	0
D	0	1
E	0	1

=

	M	F
A	2	0
B	1	0
C	1	1
D	2	1
E	1	1

Network Characteristics

- **A network consists of nodes (actors) and edges (ties).**
- **Node**
  - Attribute
  - Weight
- **Edge**
  - Attribute
  - Direction
  - Weight

Network Characteristics

– Think about it!

➔ Unweighted undirected network ≡ (Symmetric/ Asymmetric)?

➔ Unweighted directed network ≡ (Symmetric/ Asymmetric)?

➔ Weighted undirected network ≡ (Symmetric/ Asymmetric)?

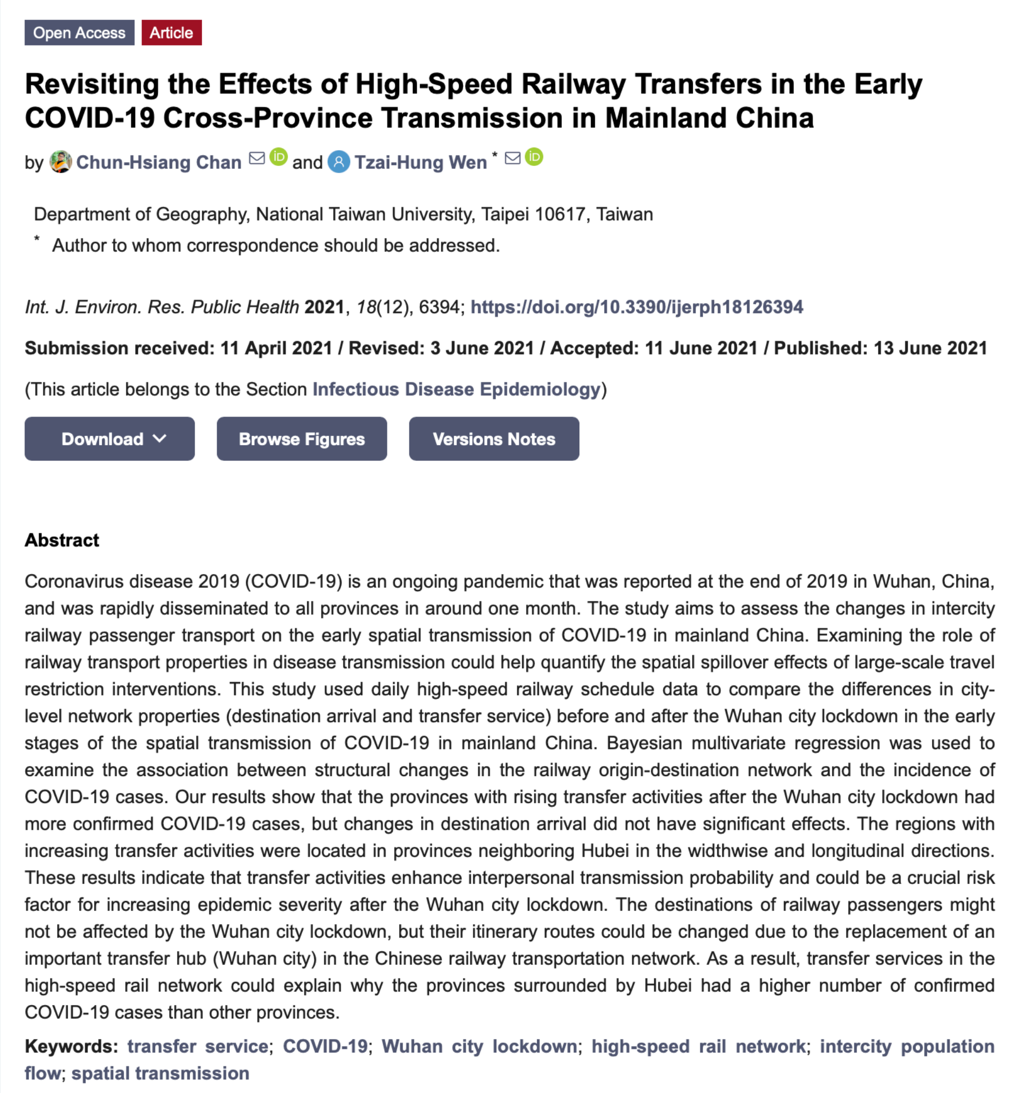
➔ Weighted directed network ≡ (Symmetric/ Asymmetric)?

Paper Reading

Chan, C. H., & Wen, T. H. (2021). Revisiting the Effects of High-Speed Railway Transfers in the Early COVID-19 Cross-Province Transmission in Mainland China. International Journal of Environmental Research and Public Health, 18(12), 6394.

Questions:

1. What is the objective of this paper?
2. What are the nodes (actors) and edges (ties) of the transportation network in this paper?
3. How did the authors formulate the transportation network?
4. What are the findings of this study?
5. If you want to achieve the same objective, how do you formulate the network?



Open Access Article

Revisiting the Effects of High-Speed Railway Transfers in the Early COVID-19 Cross-Province Transmission in Mainland China

by **Chun-Hsiang Chan** and **Tzai-Hung Wen** *

Department of Geography, National Taiwan University, Taipei 10617, Taiwan

* Author to whom correspondence should be addressed.

Int. J. Environ. Res. Public Health **2021**, *18*(12), 6394; https://doi.org/10.3390/ijerph18126394

Submission received: 11 April 2021 / Revised: 3 June 2021 / Accepted: 11 June 2021 / Published: 13 June 2021

(This article belongs to the Section Infectious Disease Epidemiology)

Download ⌄ Browse Figures Versions Notes

Abstract

Coronavirus disease 2019 (COVID-19) is an ongoing pandemic that was reported at the end of 2019 in Wuhan, China, and was rapidly disseminated to all provinces in around one month. The study aims to assess the changes in intercity railway passenger transport on the early spatial transmission of COVID-19 in mainland China. Examining the role of railway transport properties in disease transmission could help quantify the spatial spillover effects of large-scale travel restriction interventions. This study used daily high-speed railway schedule data to compare the differences in city-level network properties (destination arrival and transfer service) before and after the Wuhan city lockdown in the early stages of the spatial transmission of COVID-19 in mainland China. Bayesian multivariate regression was used to examine the association between structural changes in the railway origin-destination network and the incidence of COVID-19 cases. Our results show that the provinces with rising transfer activities after the Wuhan city lockdown had more confirmed COVID-19 cases, but changes in destination arrival did not have significant effects. The regions with increasing transfer activities were located in provinces neighboring Hubei in the widthwise and longitudinal directions. These results indicate that transfer activities enhance interpersonal transmission probability and could be a crucial risk factor for increasing epidemic severity after the Wuhan city lockdown. The destinations of railway passengers might not be affected by the Wuhan city lockdown, but their itinerary routes could be changed due to the replacement of an important transfer hub (Wuhan city) in the Chinese railway transportation network. As a result, transfer services in the high-speed rail network could explain why the provinces surrounded by Hubei had a higher number of confirmed COVID-19 cases than other provinces.

Keywords: transfer service; COVID-19; Wuhan city lockdown; high-speed rail network; intercity population flow; spatial transmission

References

- Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.
- Tsvetovat, M., & Kouznetsov, A. (2011). Social Network Analysis for Startups: Finding connections on the social web. "O'Reilly Media, Inc.".
- Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.
- https://en.wikipedia.org/wiki/Social_network

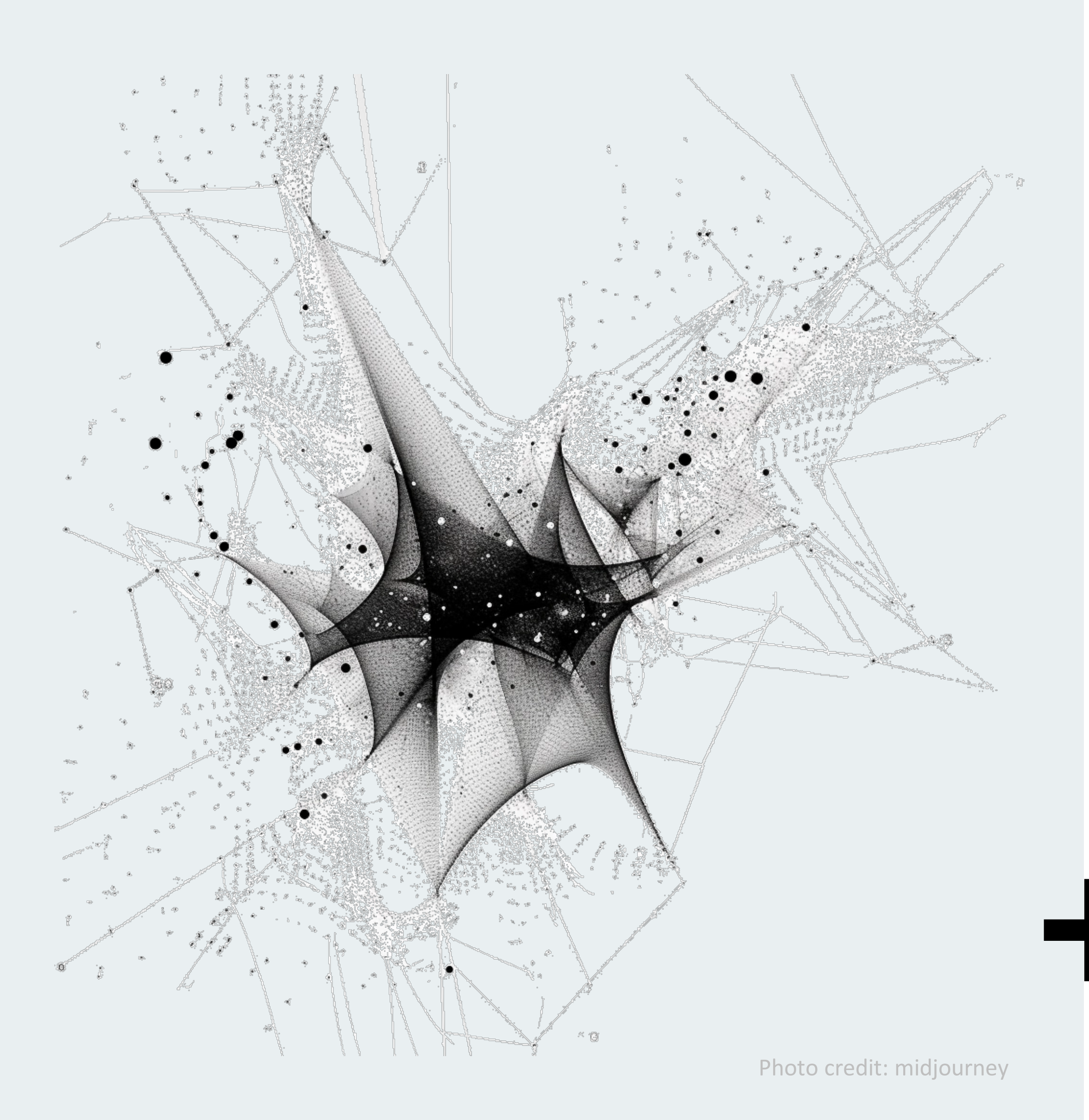


Photo credit: midjourney

Social Network Analysis

The End

Thank you for your attention!

Email: chchan@ntnu.edu.tw
Website: toodou.github.io

Photo credit: midjourney