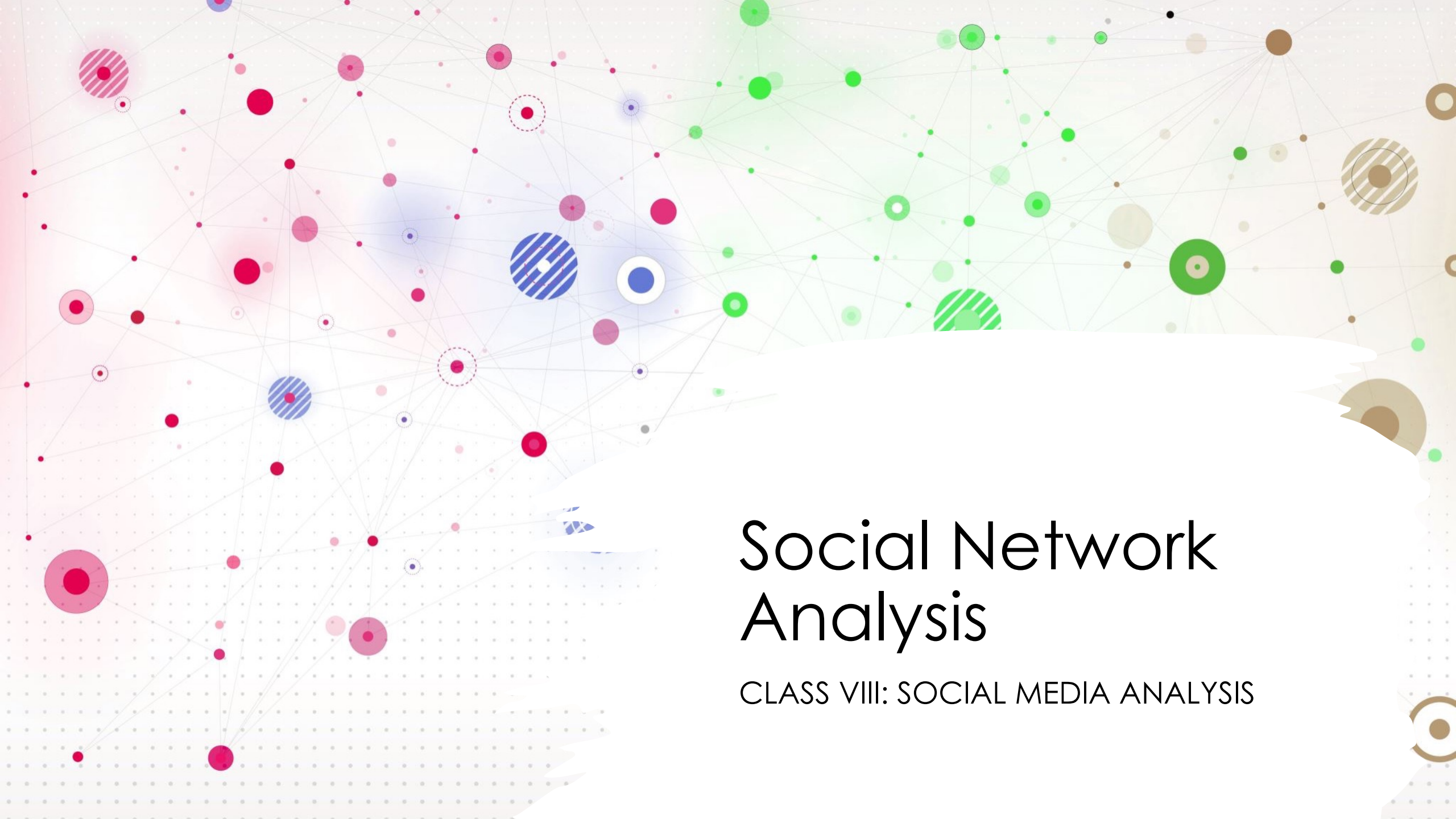


Social Network Analysis

CLASS VIII: SOCIAL MEDIA ANALYSIS

Recap

- For the last couple of days we've been looking at how to analyse Twitter data using network analysis tools.
- We'll continue with that in an R class session later. Today we'll take a brief look at how to combine text analysis with network analysis.

Your Final Assignment

- On Monday, you began outlining your final assignment for this class: a small individual project in which you'll analyse some social media data of your choice using the network analysis approaches discussed in this module.

- You identified the **data** you intend to use, which means you should also have a relatively clear idea of your **research question**.

Final Assignment Details

- For your final assignment, you will submit a short research paper, along with the code and data used for the project.
- The paper only needs to be a few pages long, but it should be structured like a full research paper:
  - Introduction (research question)
  - Methodology
  - Results
  - Conclusions

Research Project Details

- You are free to use any social media data you wish and apply any research method you wish.
  - You're not limited to the methods we studied in class; you should of course carry out a network analysis, but you may also use methods you have learned elsewhere alongside this.
- Make sure you provide a clear visualisation of the network that supports and illustrates the points you make in the paper.

Assignment Timeline

- Today (Wednesday):
  - Begin working on scraping / downloading the data you need for your project and write a short report (a couple of paragraphs) indicating your progress, any problems you have encountered, and any changes / updates to your research plan.
- Tomorrow (Thursday):
  - Prepare a presentation of your full research plan, outlining briefly your question and hypothesis, and the research method you plan to use.
  - Each student will present this research plan in-class on Friday. Your presentation should fully explain what you intend to do and why. Slides are not required but you may prepare some if you prefer. (Students who cannot attend class may submit a written plan instead.)


Node Similarity

Node Similarity

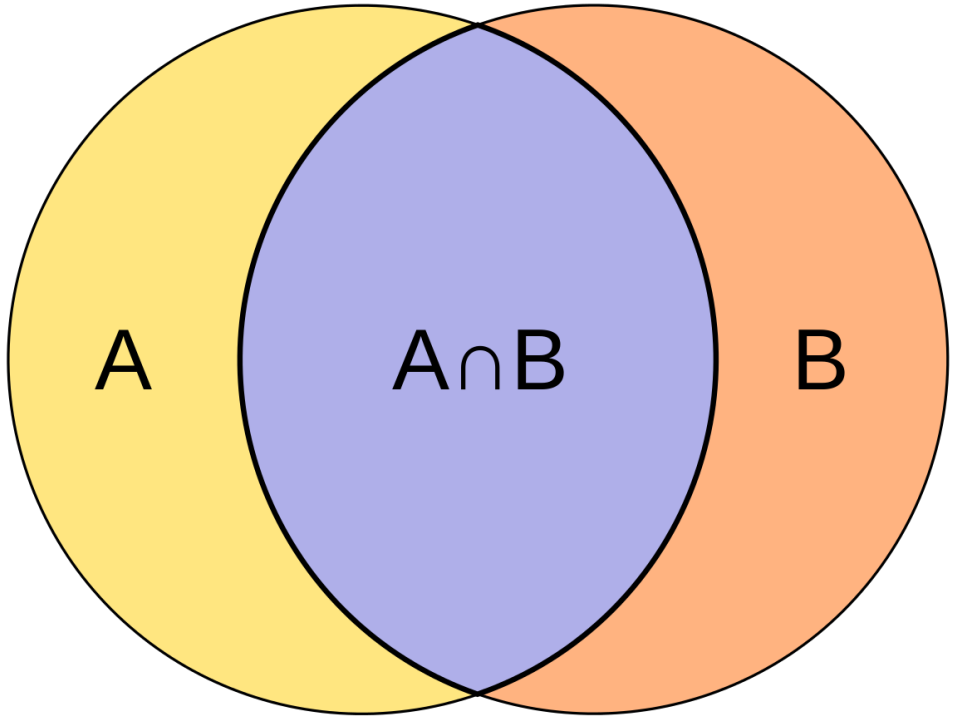
- So far, we've analysed networks in which the edges represented interactions between two nodes – retweets, friendships, shared scenes in a TV show, etc.
- However, it's also possible for edges to represent similarity between two nodes.

Twitter User Similarity Measures

- *For example:*
- If you download the Friend list of some Twitter users (all the people they follow), you can calculate the overlap between those friend lists to find a measurement of similarity.
- This is usually calculated on a pairwise basis as the Jaccard Index of the two users.

Jaccard Index

- The Jaccard Index is the Intersection over the Union of two sets.
- When inverted (1/n), it is called the Jaccard Distance between two sets.



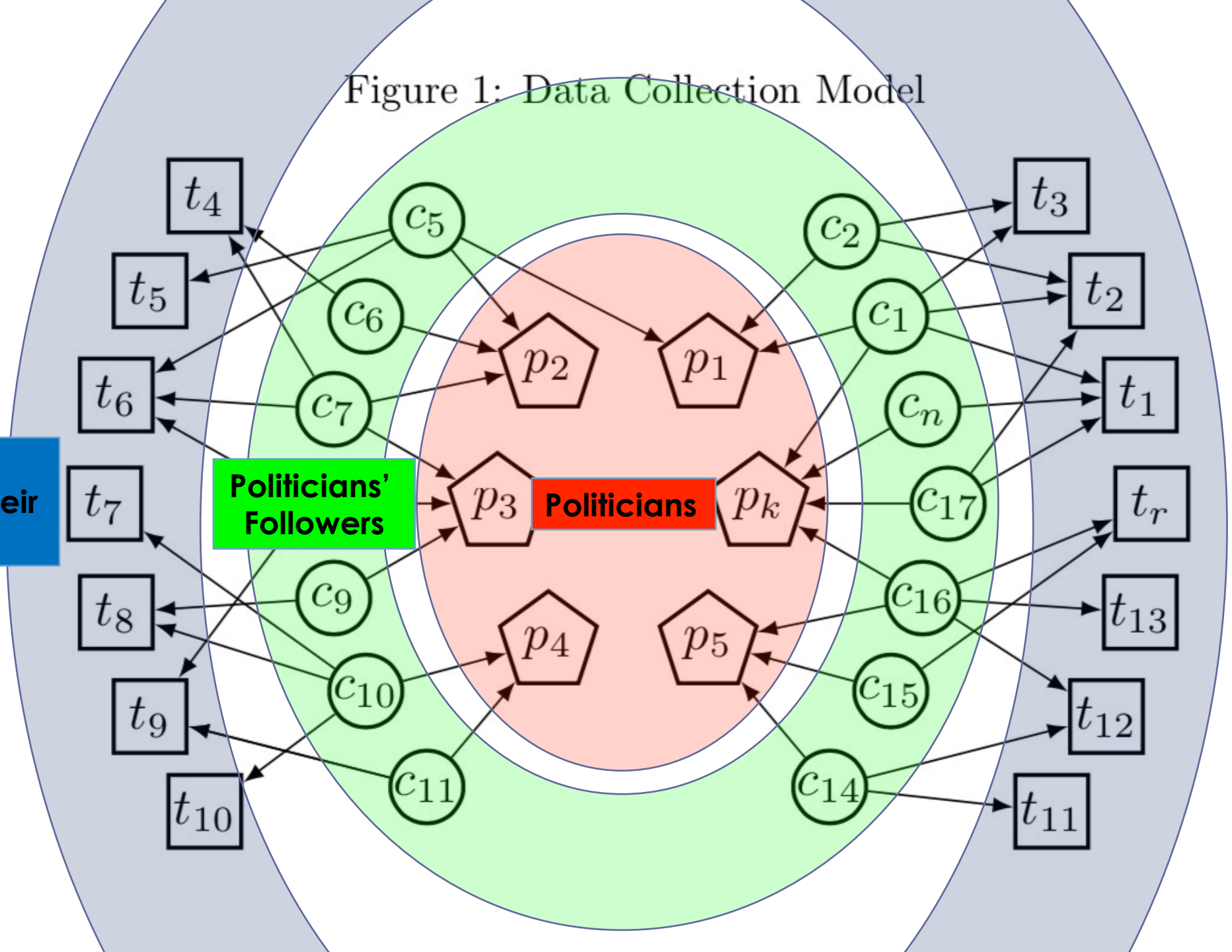
Why is this useful?

- Calculating the pairwise similarity of nodes allows us to group Twitter users together according to how similar their social media environment is.
- Users with high Jaccard Indices will have more similar information exposure.
- This is of course self-selected; users' Jaccard Indices are a measure of how similar their following choices on Twitter were.

An example…

I've mentioned this project before: here are some slides and results from the study of social polarisation on Twitter which I presented (with my co-author Stefano Camatarri) at the APSA Conference last September.

Figure 1: Data Collection Model



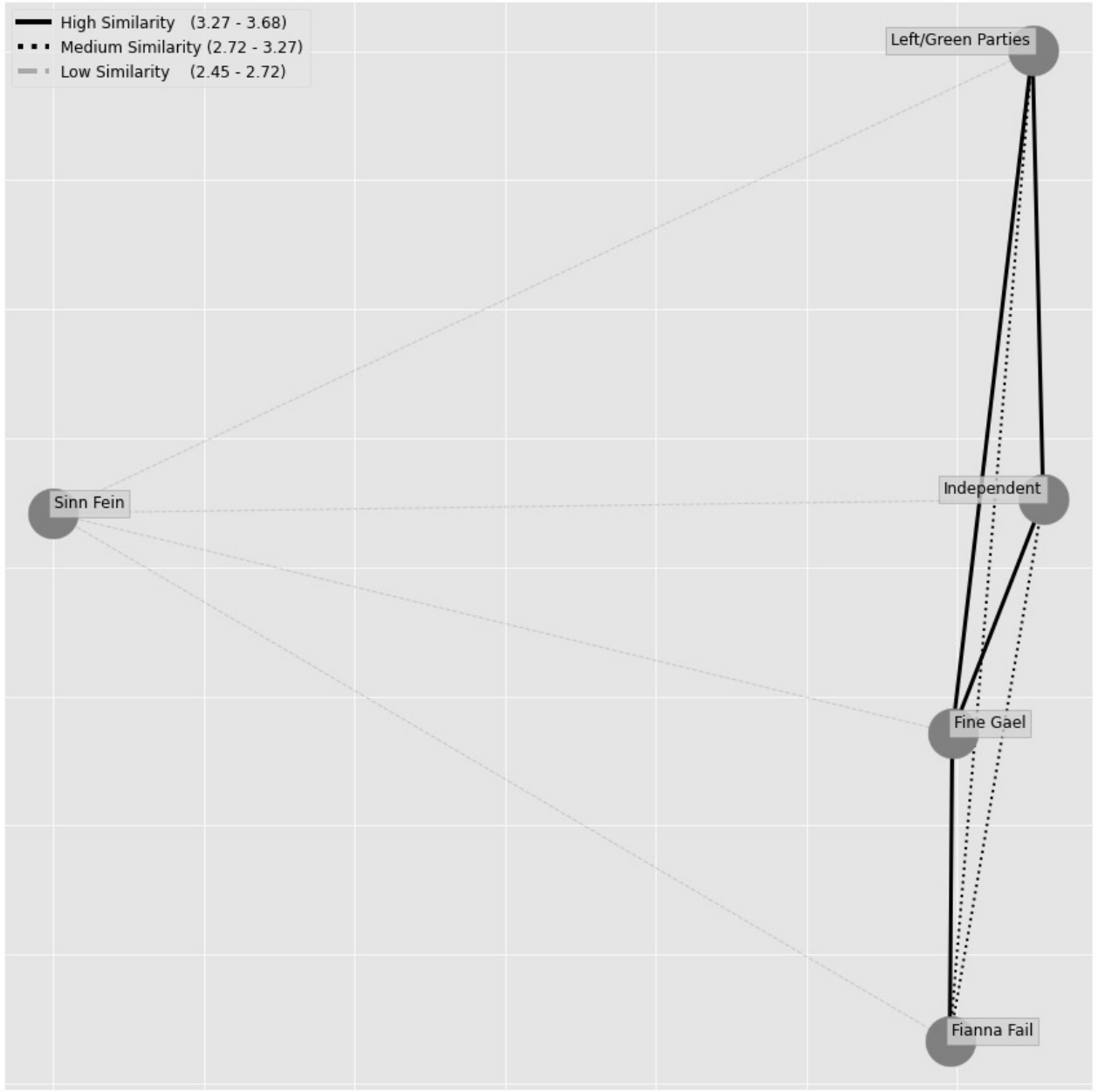
Cases in the Initial Study

	Ireland	Belgium	Italy	Japan
Population	5m	11m	60m	120m
GDP Per Capita	$61,170	$50,760	$41,430	$43,490
GINI	32.9	27.4	35.9	32.8
Existing Cleavage?	**No**	**Yes:** Socio-cultural cleavage between Dutch- and French-speaking populations.	**Yes:** Economic cleavage between industrial/wealthy north and agricultural/less developed south.	**No**
Language(s)	English (Irish minority)	Dutch, French	Italian	Japanese

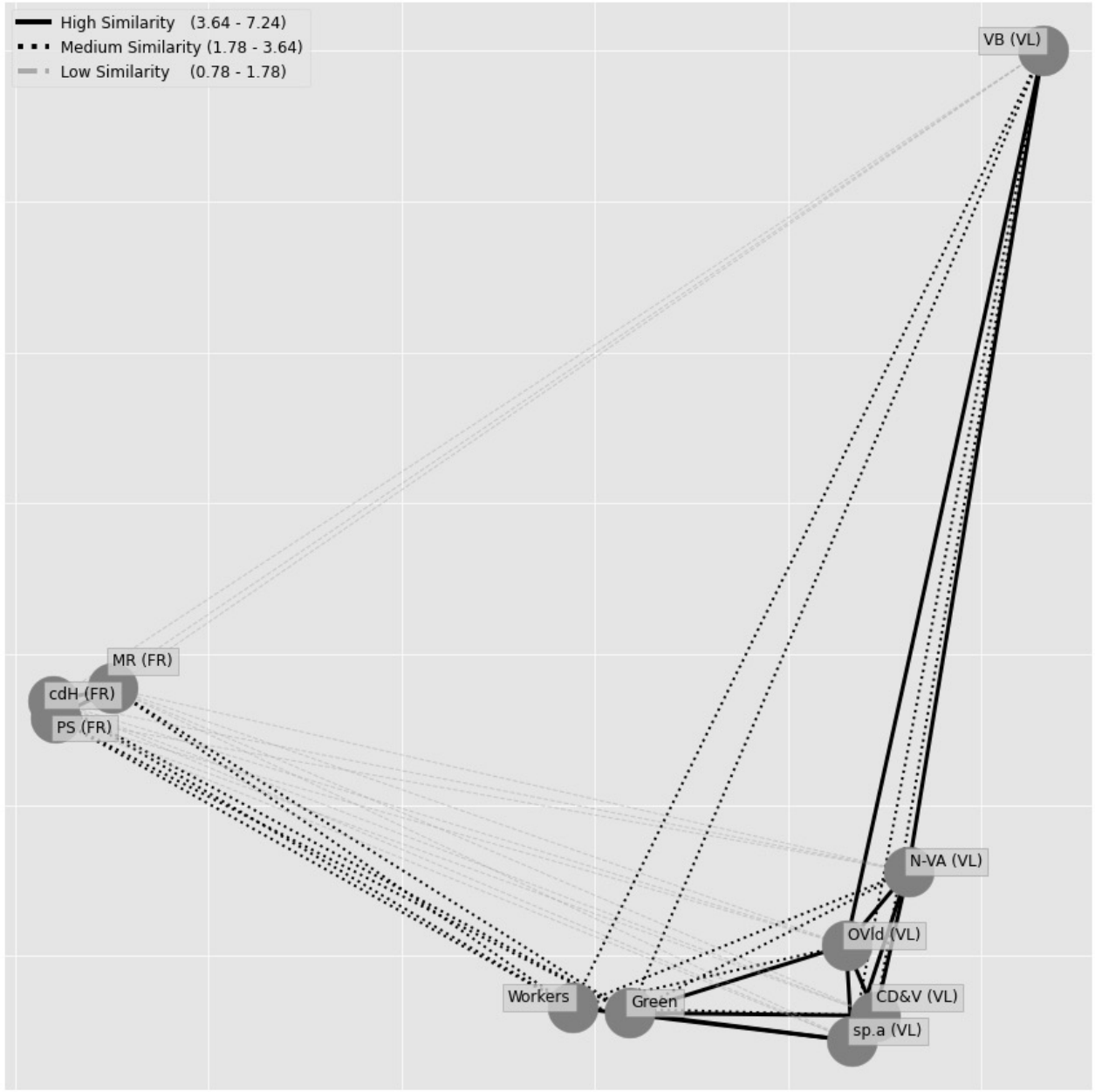
Method

- Scrape Twitter followers of every politician. Randomly sample 10,000 users who follow at least five of them.
- Assign a political leaning based on which party they follow most. Scrape the list of other accounts which they also follow.
- Calculate Jaccard Index of party followers' non-political following lists to calculate spatial distance (network edges).

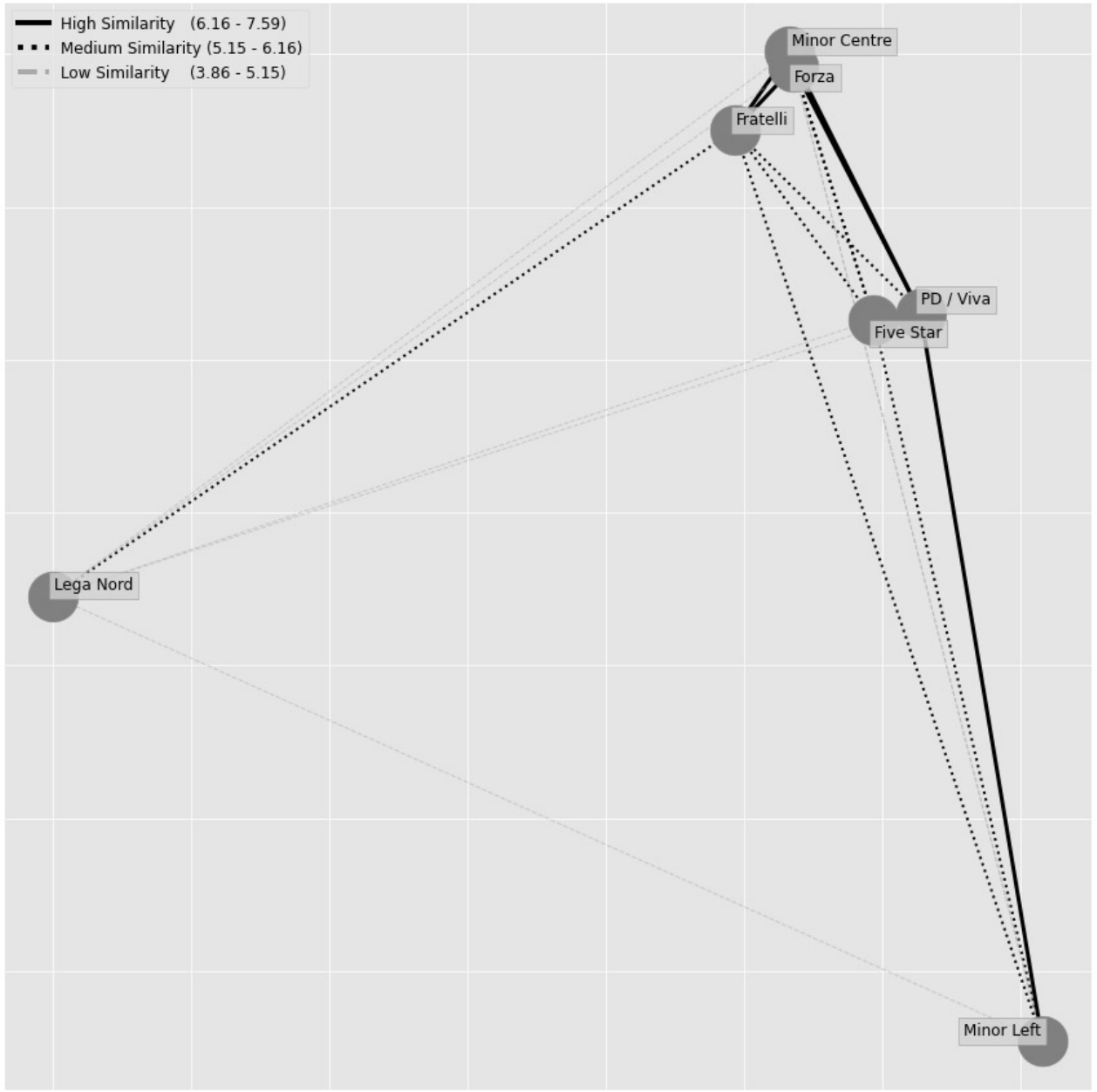
Ireland



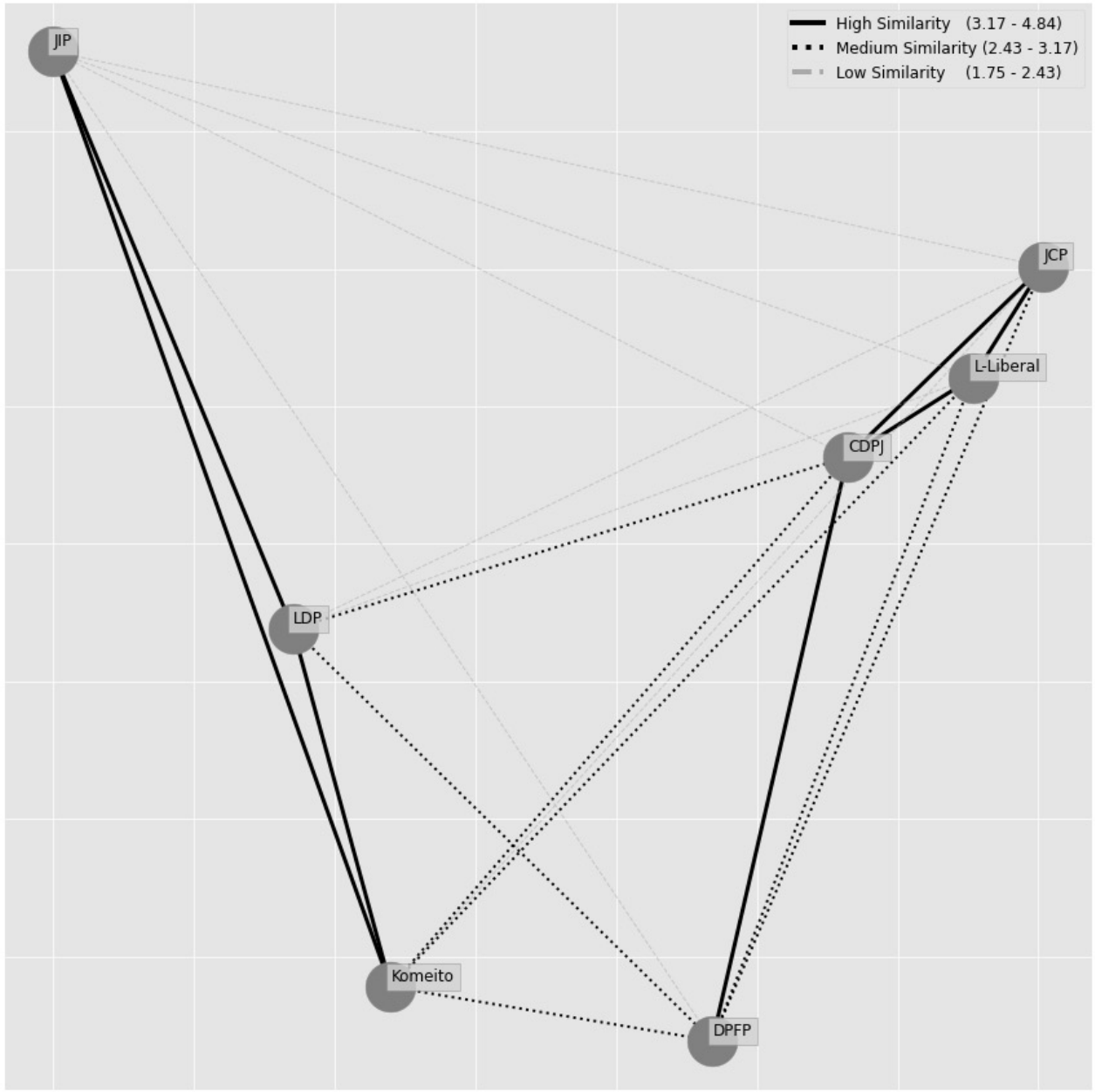
Belgium



Italy


High Similarity (6.16 - 7.59)
Medium Similarity (5.15 - 6.16)
Low Similarity (3.86 - 5.15)
Minor Centre
Forza
Fratelli
PD / Viva
Five Star
Lega Nord
Minor Left

Japan


High Similarity (3.17 - 4.84)
Medium Similarity (2.43 - 3.17)
Low Similarity (1.75 - 2.43)
JIP
JCP
L-Liberal
CDPJ
LDP
Komeito
DPFP

- **Ireland** and **Japan** – both countries with no theorized existing cleavage – show low levels of non-political polarization.
- **Belgium**'s polarization was the highest, especially between VB and the French Parties.
  - In comparative context, the gap between VB and other Flemish parties turns out to be very minor.
- In **Italy**, Lega Nord is strongly polarized from other parties.