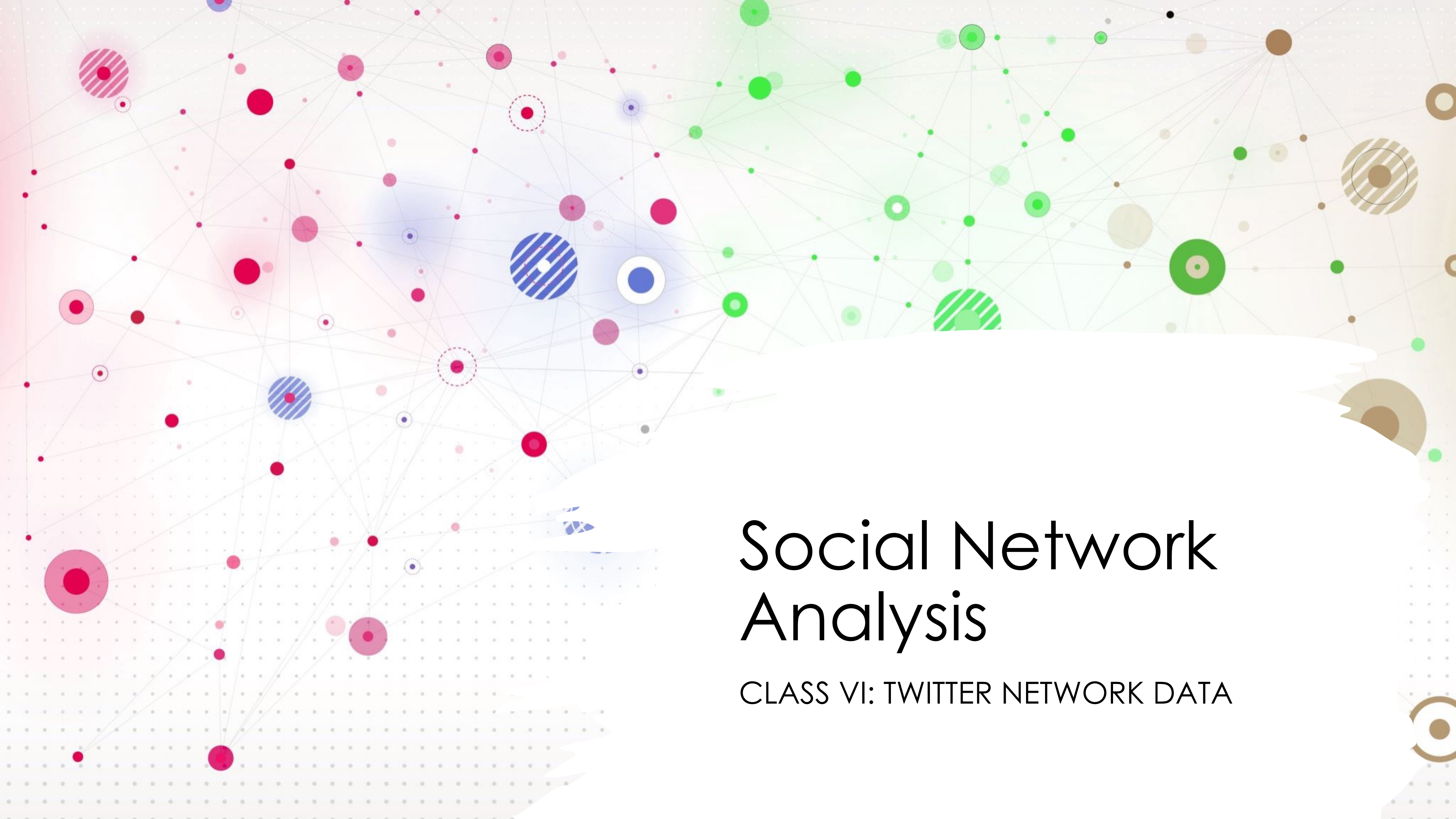
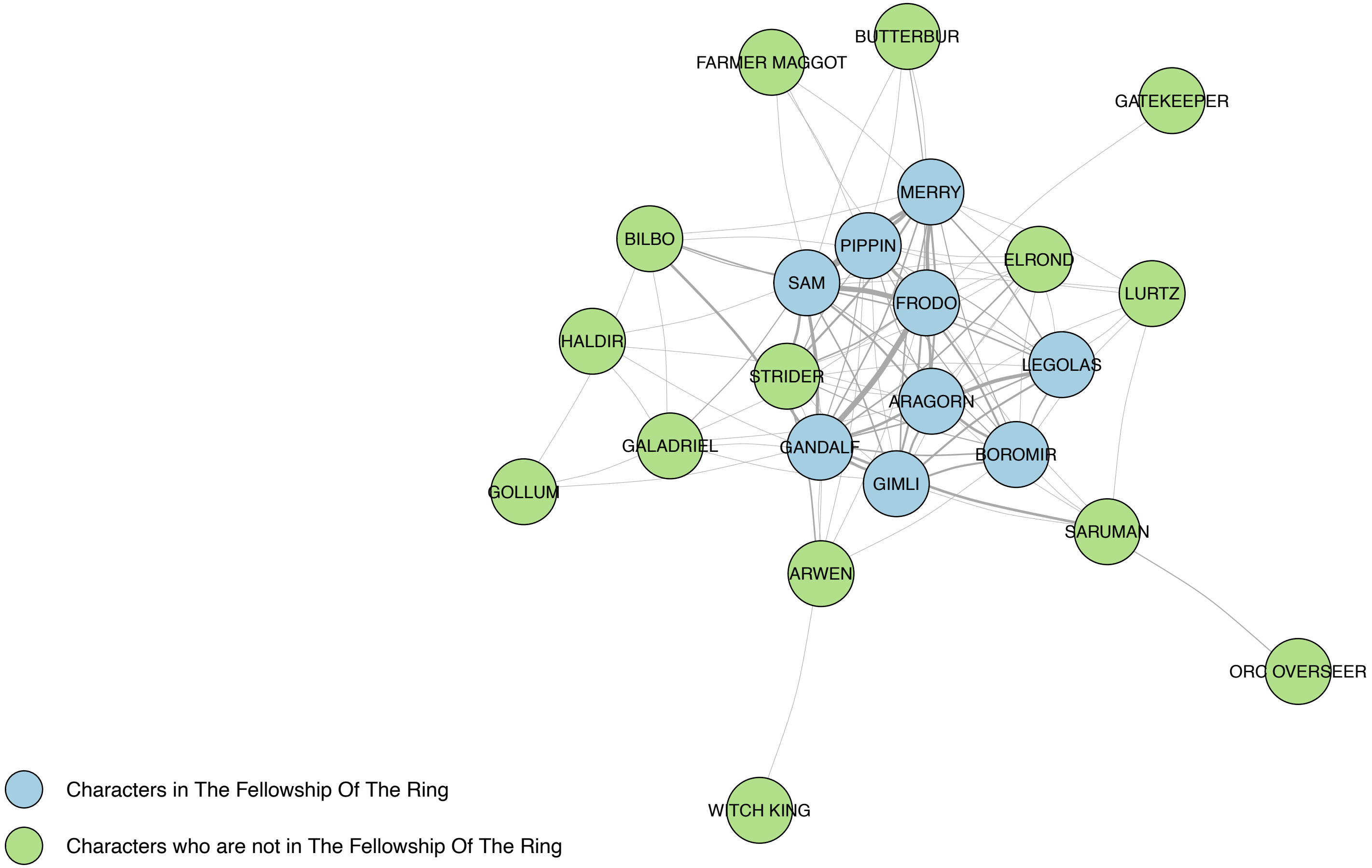


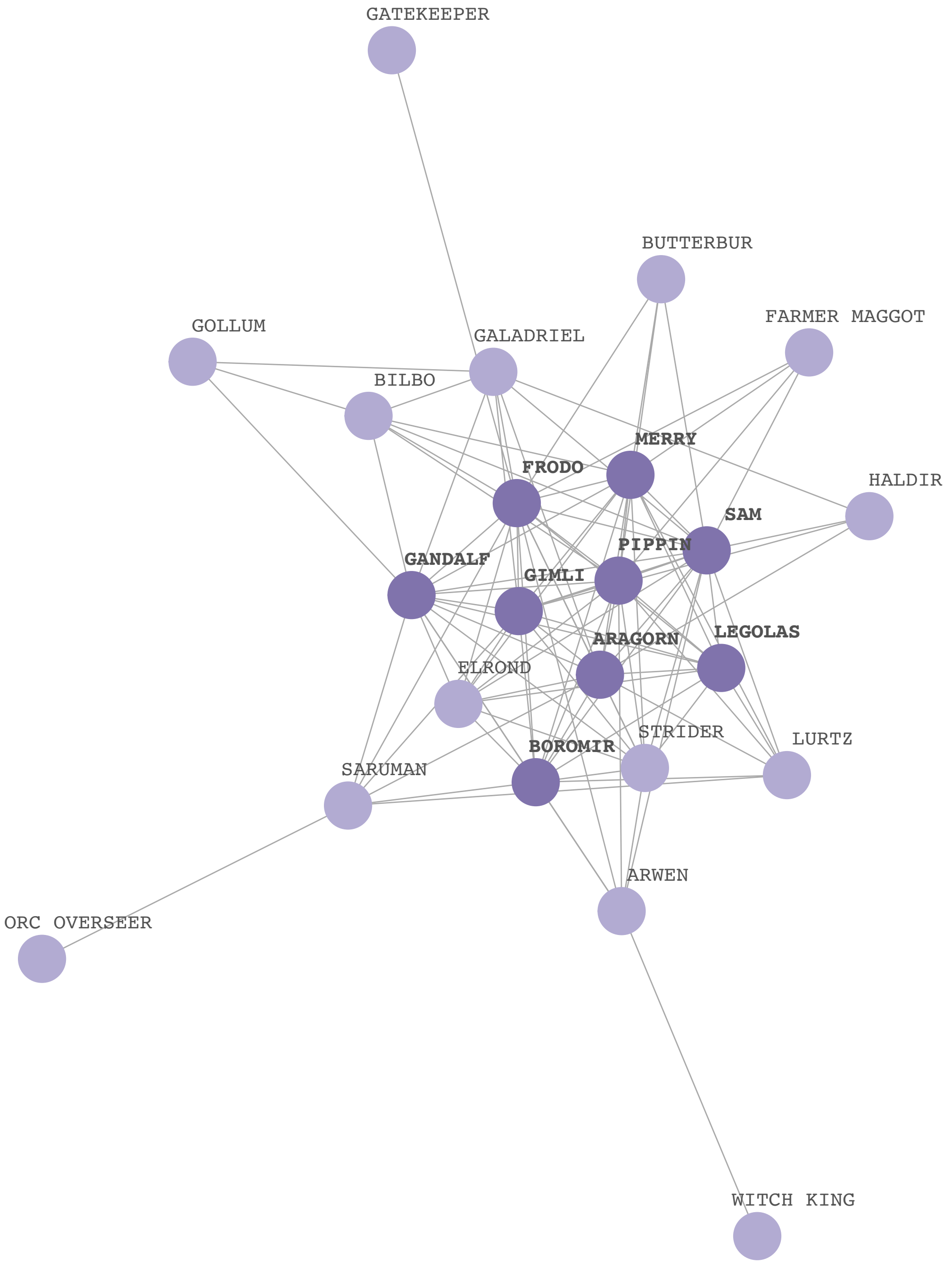
Social Network Analysis

CLASS VI: TWITTER NETWORK DATA

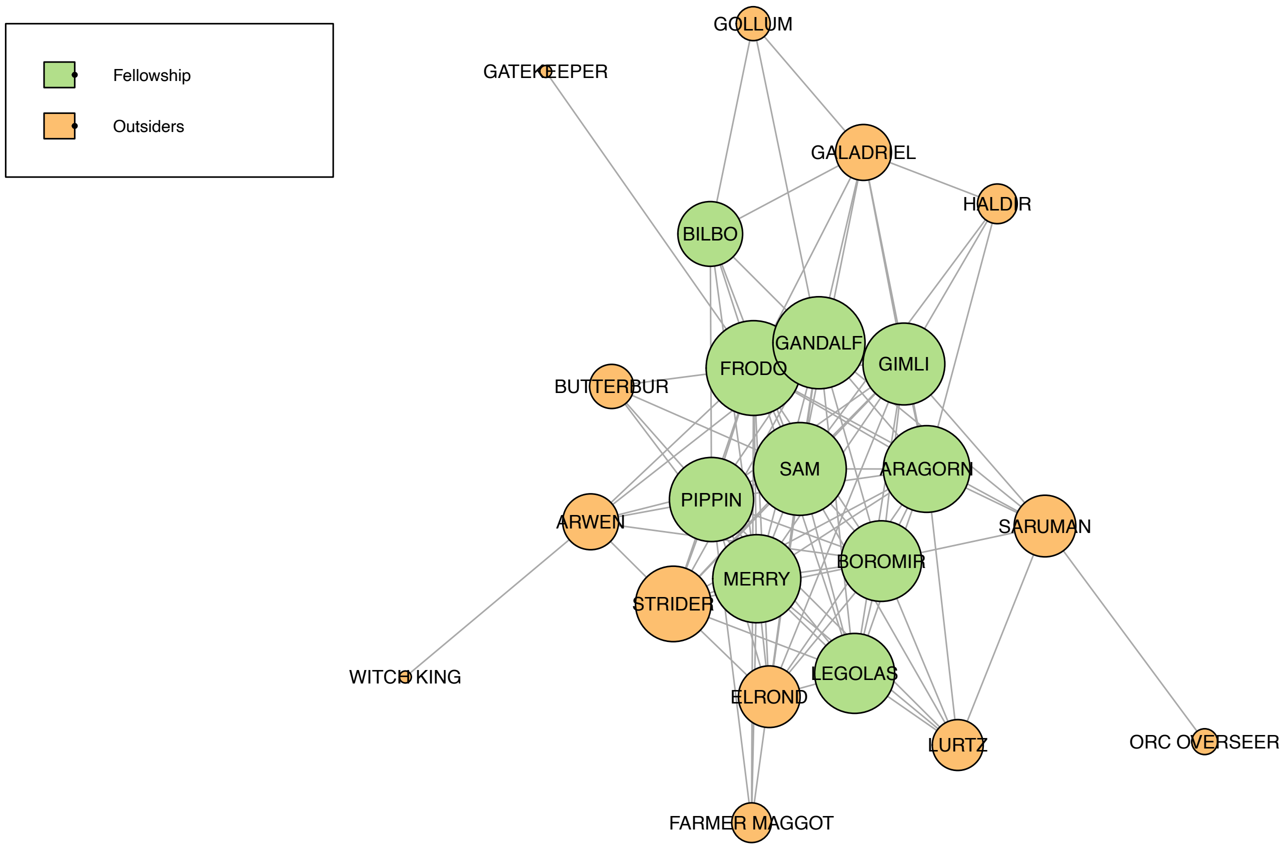
Recap

- Last week, we looked in-depth at various methods and techniques for studying small- and medium-scale networks.
- This week, we're moving on to looking at how to access and analyse one of the most ubiquitous – but difficult! – types of network data, digital data from social media.

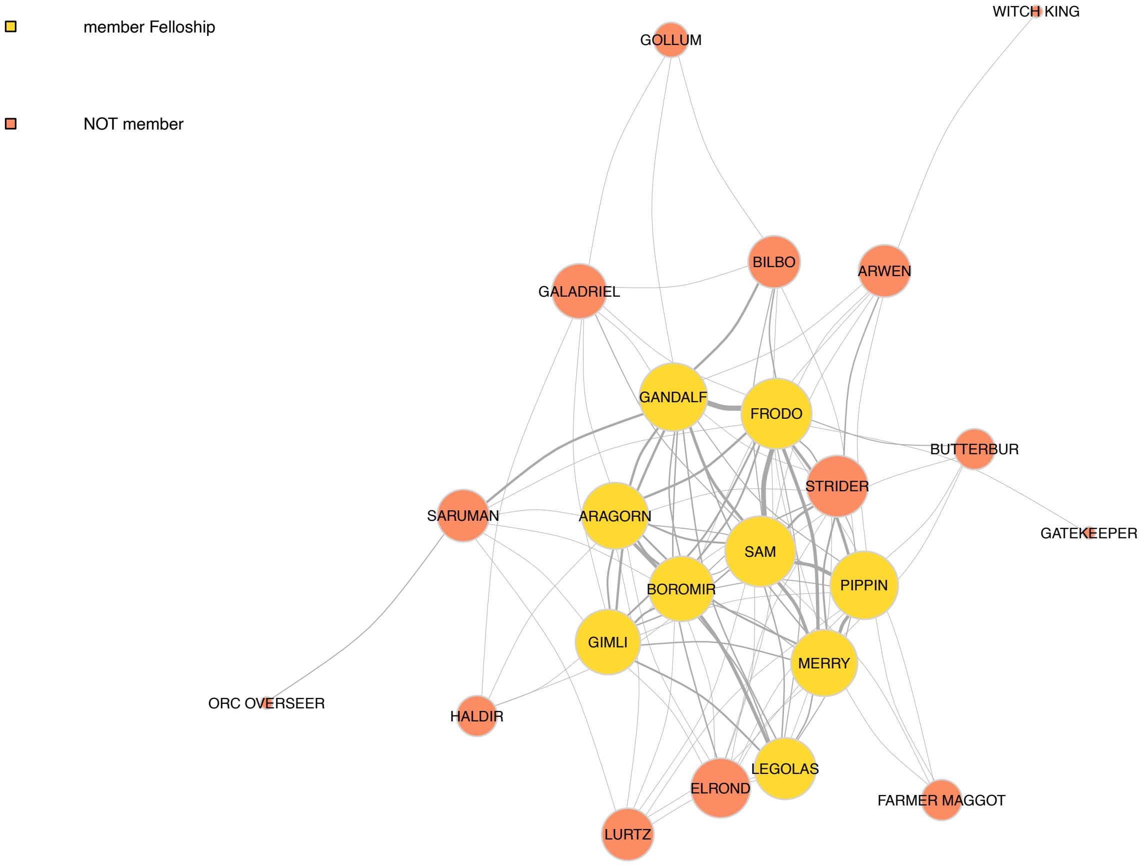

BUTTERBUR
FARMER MAGGOT
GATEKEEPER
MERRY
PIPPIN
BILBO
ELROND
SAM
FRODO
LURTZ
HALDIR
LEGOLAS
STRIDER
ARAGORN
GALADRIEL
GANDALF
BOROMIR
GOLLUM
GIMLI
SARUMAN
ARWEN
ORC OVERSEER
WITCH KING
Characters in The Fellowship Of The Ring
Characters who are not in The Fellowship Of The Ring

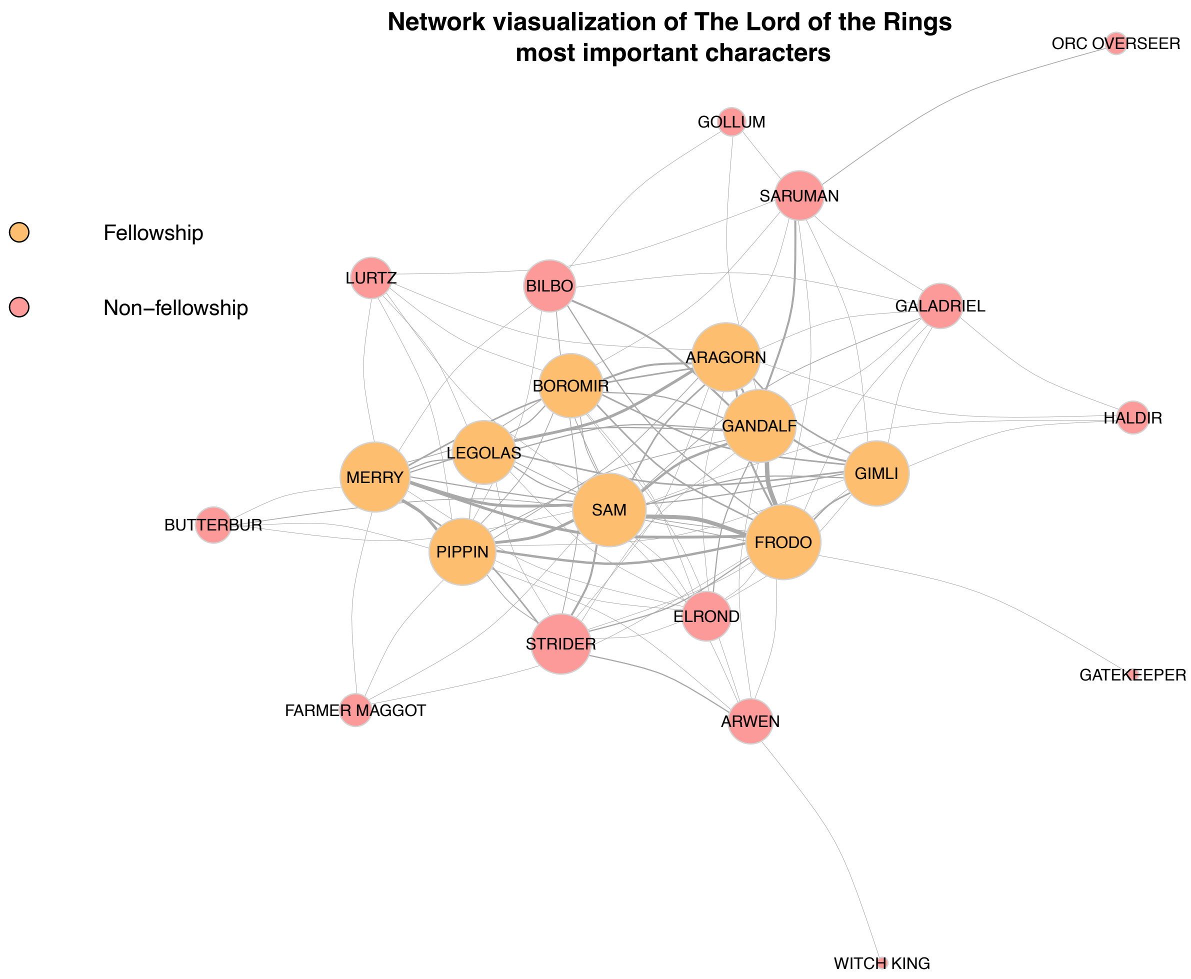

GATEKEEPER
BUTTERBUR
GOLLUM
GALADRIEL
FARMER MAGGOT
BILBO
MERRY
FRODO
HALDIR
SAM
PIPPIN
GANDALF
GIMLI
LEGOLAS
ARAGORN
ELROND
BOROMIR
STRIDER
LURTZ
SARUMAN
ARWEN
ORC OVERSEER
WITCH KING

Lord of The Rings Network Analysis: the Fellowship of the Ring



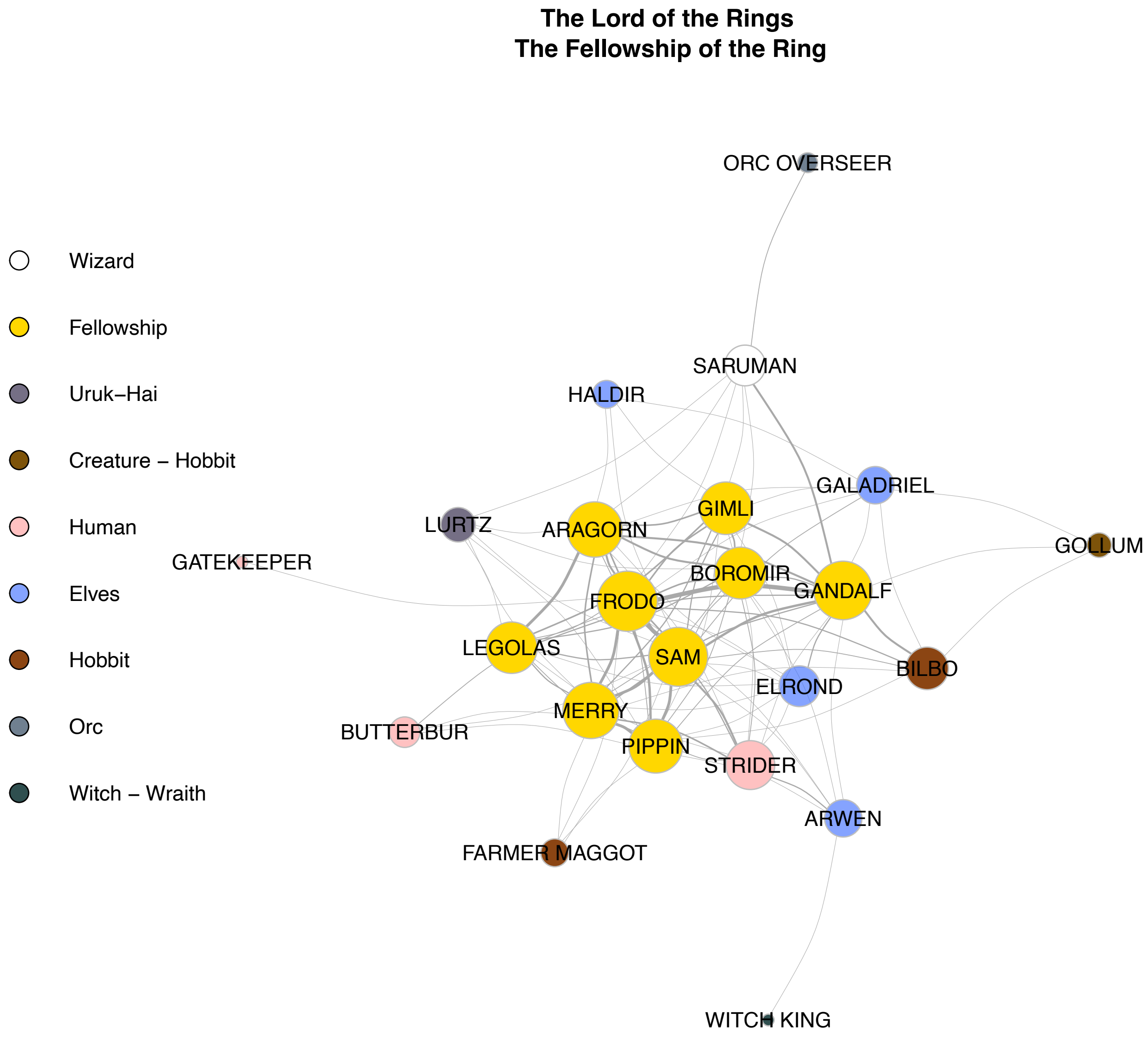
Lord of Rings network analysis





Network viasualization of The Lord of the Rings most important characters

Size of the nodes proportional to the number of scenes in which the characters appear in together



The Lord of the Rings
The Fellowship of the Ring

Twitter Data

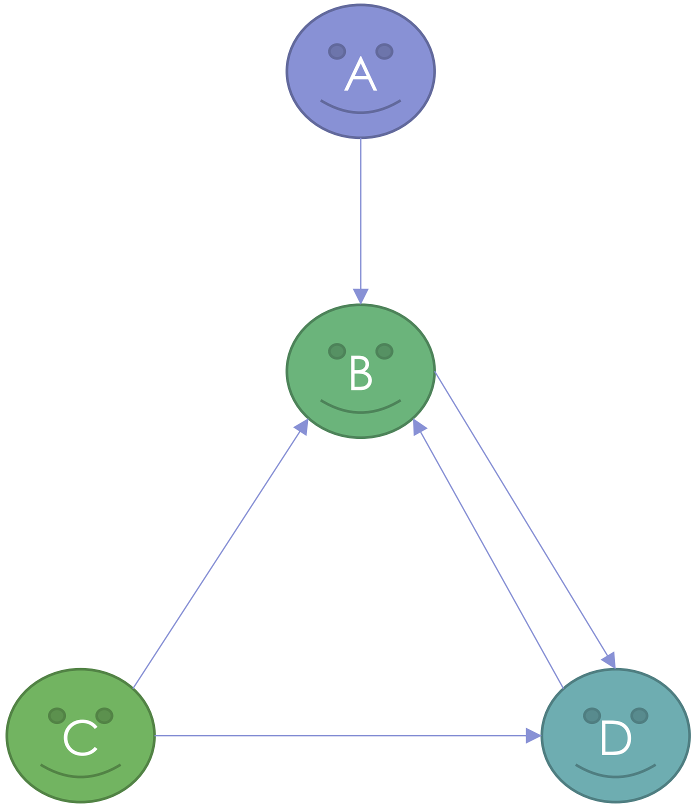
- We’re going to focus in this module on data from Twitter.
- You’ve worked with Twitter data previously so you (presumably) have the API access keys etc. set up for this purpose.
- More importantly, Twitter is a mostly **open** network; its data is available to researchers (with restrictions), unlike Facebook / Instagram / Snapchat.

Networks on Twitter

- Twitter data contains two basic types of network – **follow networks** and **retweet networks**.
- These networks are related to one another in some ways.
  - Follower networks are the basis for retweet networks, since you won’t actually see a tweet unless you follow someone who retweets it!
- However, they’re dramatically different in other key ways, and you need to choose the right one to study for your research.

Follower Networks

- Follower networks are the networks created by Twitter users' **following** decisions.
- Each user **follows** a certain number of other users and is in turn **followed** by a certain number of users.
- This is a **directed** network; these relationships do not have to be reciprocal.

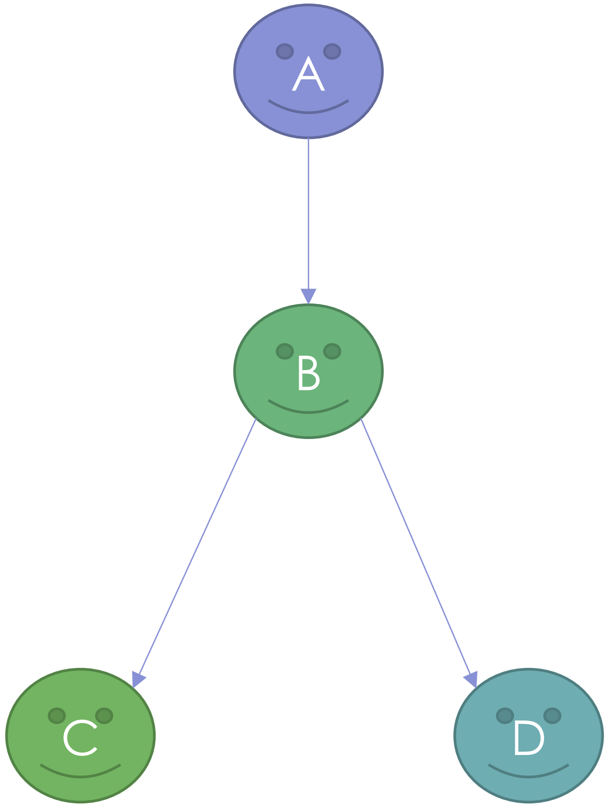


Follower Networks

- Follower networks **evolve slowly**.
- Twitter users generally do not make follow / unfollow decisions very regularly; their following network will generally change only a little bit from month to month.
- Moreover, they may choose to **mute** users they become dissatisfied with rather than **unfollowing** them, which further slows the pace of network change.

Retweet Networks

- Retweet networks are networks created by users sharing each other's posts.
- They are also **directional** – a retweet is a one-way interaction.
- They tend to be **hierarchical**, meaning that the network spreads outwards from a core.



Retweet Networks

- Retweet networks are **rapidly-forming** and **transient**.
- The network can form and evolve extremely quickly in response to an event.
- The network disappears just as quickly!
  - Retweeting a user does not create any permanent network link.
  - Within a few days or even just hours, the network effect of a retweet fades and eventually disappears.

Downloading Twitter Network Data

- Both types of network can be accessed through the Twitter API.
- (You probably have this set up from a previous class – you'll need your access keys and tokens from the **dev.twitter.com** site.)
- We'll be using the **rtweet** library to access the API.

Follower Networks

Downloading Follower Networks

- Following information is accessed through the **User** object in the Twitter API.
- Each User object has two fields indicating how many accounts they **follow** and how many they are **followed by:**
  - **followers_count**
  - **friends_count**
    - (Twitter calls people you follow “friends” – don’t be confused by this notation, these are still unidirectional relationships, not reciprocal ones!)

Downloading Follower Networks

- The User object in TwitteR also has a set of methods for accessing the full list of followers / friends:

Followers	Friends
get_followers()	get_friends()

- These commands return a list of Twitter IDs for the accounts; to access the full user object with all its data, we then need to run **lookup_users**() for these IDs.

Twitter Rate Limits

- It may be tempting to download the full User object for all the accounts, not just their Twitter ID – but unless you have a genuine need for information in that User object, **don't do this**.
- Accessing User objects and follower / friend lists is subject to very strict **rate limits** on Twitter.
  - It can take **days** of running your code to download the followers for a single popular Twitter account – and it's **much slower** if you're downloading all the full-size User objects.

An example…

- Two years ago we downloaded the **follower list** for every politician in the Italian parliament who has a Twitter account.
- This process took **six full days** to complete and resulted in a database of 19.2 million entries (i.e. the resulting network had 19.2 million edges).
  - That’s even though only two politicians (Renzi and Salvini) had more than a million followers.
  - Imagine trying to do the same for US politicians – or for major celebrities!

Retweet Networks

Downloading Retweet Networks

- Generally speaking, **retweet networks** are constructed by:
  - Downloading a large volume of tweets based on a certain criteria;
  - Finding the retweets in that data set and creating a network edge from the retweeter to the person they retweeted.

Tweet Search Criteria

- The nature of your retweet network depends on the criteria you use to search Twitter and gather tweets.
- You can specify keywords and then filter your data to observe a <u>single tweet</u> to see how it propagated over a network;
- Or you can focus on broader keywords to see how information on a topic more generally has moved around.

Downloading Retweet Networks

- Generally speaking it’s easier to access data for retweet networks, as the **rate limits** for downloading tweets are less restrictive than for followers/friends.
- However, the usual rules regarding the Twitter API apply; without paying or getting a special exemption, you cannot search for data more than 7 days old, and the data you access will be a sample, not the full data set.

A quick note on time series analysis…

- Follower / Friend networks don’t have any **temporal** data – i.e. you have no way of knowing when the follower/friend relationship was created. This makes it generally impossible to model their development over time.
- Retweets have a **timestamp**, so it’s easy to filter your network by timestamp and show how the information moved between different clusters over time.

Okay, let's go over and look at this in R.