Class 3: Measuring Properties of Groups and Networks

ISOM 673:
Social Network Analytics

Class 3

- Demetrius Lewis, PhD
- Emory University
- October 27, 2017

Class Purpose

- The purpose of this session is to teach you how to:

  - **Develop concepts of network subgroups and cliques**
  - **Determine how clusters in the network might be found**
  - **Use the network as a whole to determine relationships between groups**
  - **Learn about small worlds and effective distances in networks**
  - **Examine whether a network has core and periphery features**
  - **Understand how to take group and network measurements and draw conclusions from network data**

Today's Class Content

- Brokerage and Closure
- Identifying Cliques in a Network
- Defining Clusters with Distance and Relationship Correlation
- Hierarchical, k-*means*/medoid, and Model-based Clustering
- Clusters versus Core-Periphery
- Core Periphery Structure in Venture Capital
- Small Worlds in the Actor Network Using the Oracle of Bacon and Neo4j
- Wrap-up

Brokerage and Closure

Benefits to Brokerage Can be Summarized along Information and Control

Information

- ***Access:*** getting information & knowing who can use it
- ***Timing***: getting opportunities to act early on fresh information
- ***Referrals:*** since you can only (usually) be in one place at one one time, maximize exposure to your skills and brand through other people's word of mouth

Control

- ***Gatekeeper:*** determine who gets access to information and resources
- ***Matchmaker:*** determine who gets to be in a relationship

"Union" strategies bring together disconnected individuals. These actions fill in the structural hole, but they increase reputation and provide more opportunities for referrals and reciprocity later on.

Closure Provides Trust and Reputation, and Increases Bandwidth

Trust and Reputation

- ***Trust:*** Helps others commit to plans or exchanges without formal guarantees
- ***Reputation:*** Repeated interactions between group members removes the need for formal vetting, acts as social "grease" that facilitates more exchanges to occur

Bandwidth

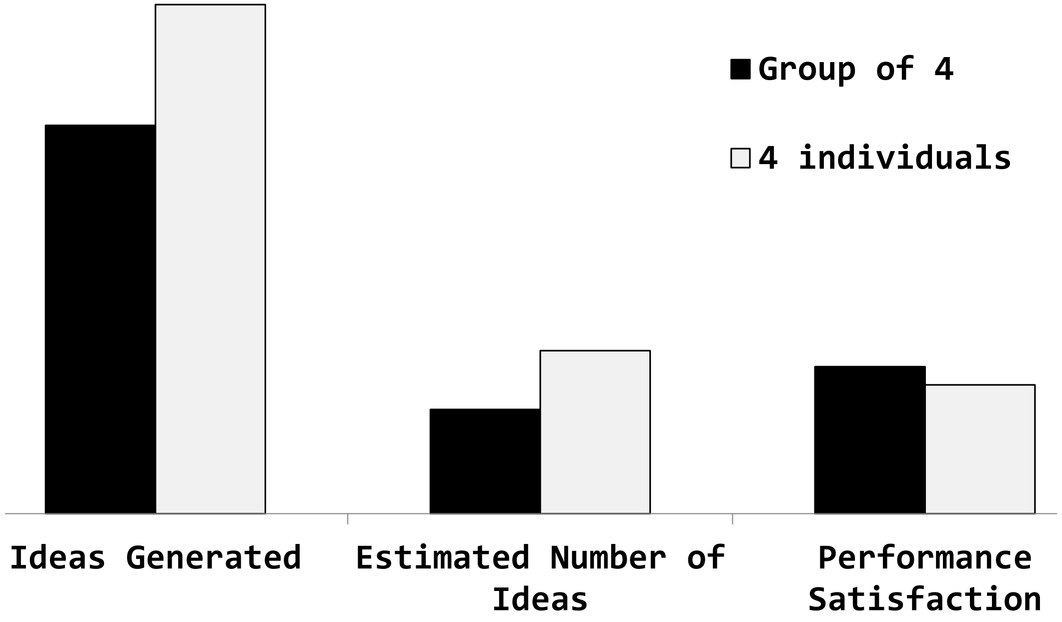
- ***Information quality:*** Reputational sanctions for providing bad information
- ***Information sharing:*** Reputational sanctions for holding onto information or acting opportunistically

However, benefits to bandwidth assume that information is accurate and unbiased. This is not always the case.

Groupthink and Norms for Relationship Management May Decrease Usefulness of Information

Groupthink

Illusion of collective success



Groups may become "echo chambers" of positive sentiment. We often solicit information from trusted contacts with similar views of others, reinforcing our prior beliefs—e.g., "I'll ask my friend." We also often politely provide information to others that is consistent with what their predispositions, filtering out contrary evidence—e.g., "I thought so, too."

Source: Stroebe, W., Diehl, M., & Abakoumkin, G. The illusion of group effectivity. *Personality and Social Psychology Bulletin* 18, no. 5 (1992): 643-650.

Identifying Cliques in a Network

In Network Analysis, we a

How to distinguish between…

- Do common friends of an actor end up being friends?
- What collections of proteins in a cell work closely together?
- Does Web page structure separate itself relative to its content?
- What portion of the Internet topology constitutes a ‘backbone’?

Cohesive Subgroups in a Network

"Actors connected via dense, directed, reciprocated, relation"

- Desirable properties include: density, reachability, familiarity, robustness
- We can define cliques along the properties above

Graduation from elementary school children to Florentine families

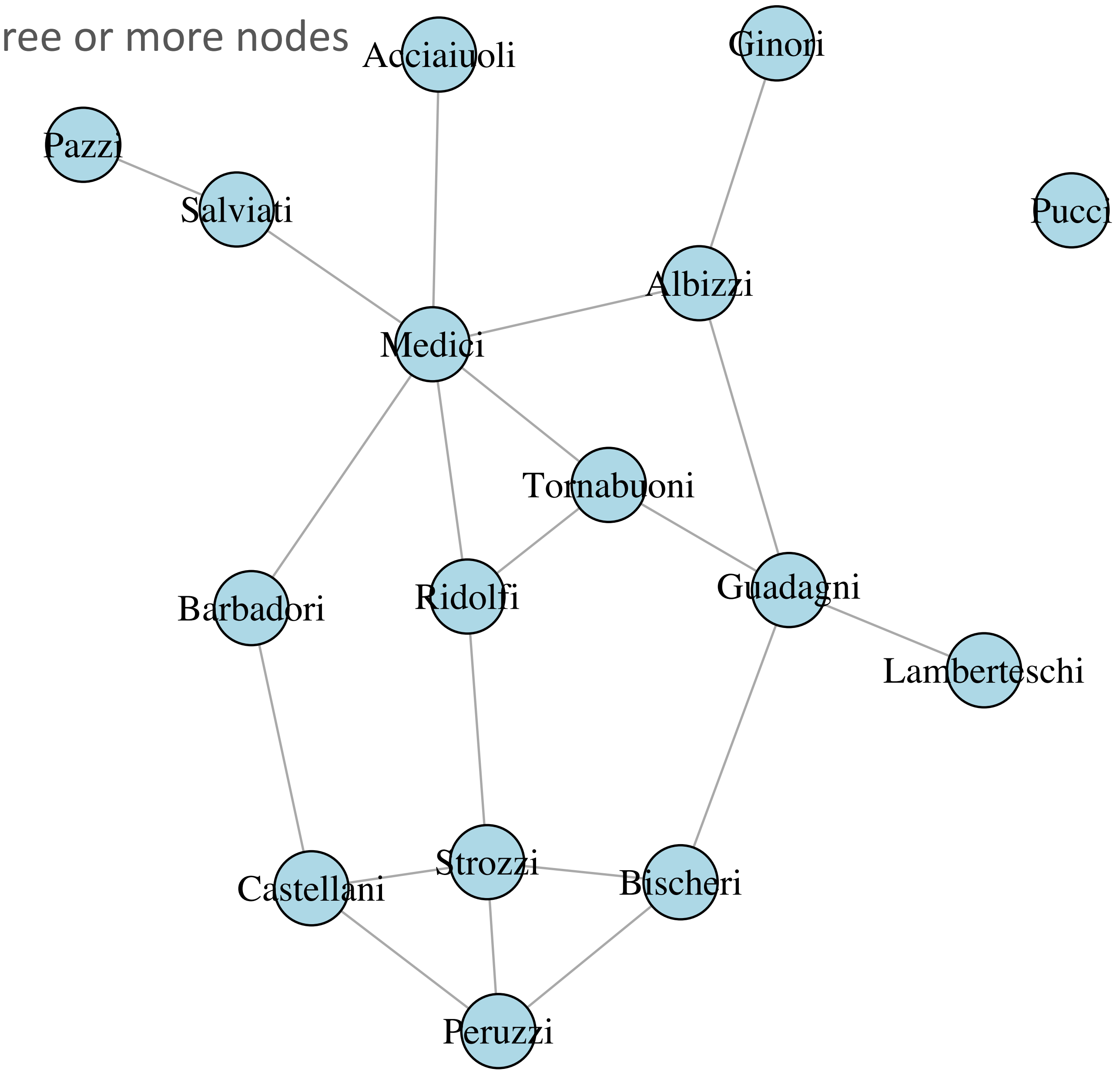
Marriage Network of Families in Florence During Renaissance Period

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Pucci	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Albizzi	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
Barbadori	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
Bischeri	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0
Castellani	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
Ginori	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Guadagni	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1
Lambertes chi	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Medici	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1
Pazzi	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Peruzzi	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0
Pucci	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ridolfi	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
Salviati	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Strozzi	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0
Tornabuon i	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0

Subgroups Based on Complete Density

Define a **clique** as a **maximal complete** subgraph of three or more nodes

- Use the minimum of 3 to exclude dyads being counted as cliques
- "Maximal complete" means that all nodes in the clique choose each other (complete), and no other node in the network that also chooses and is chosen by all the members (maximal)

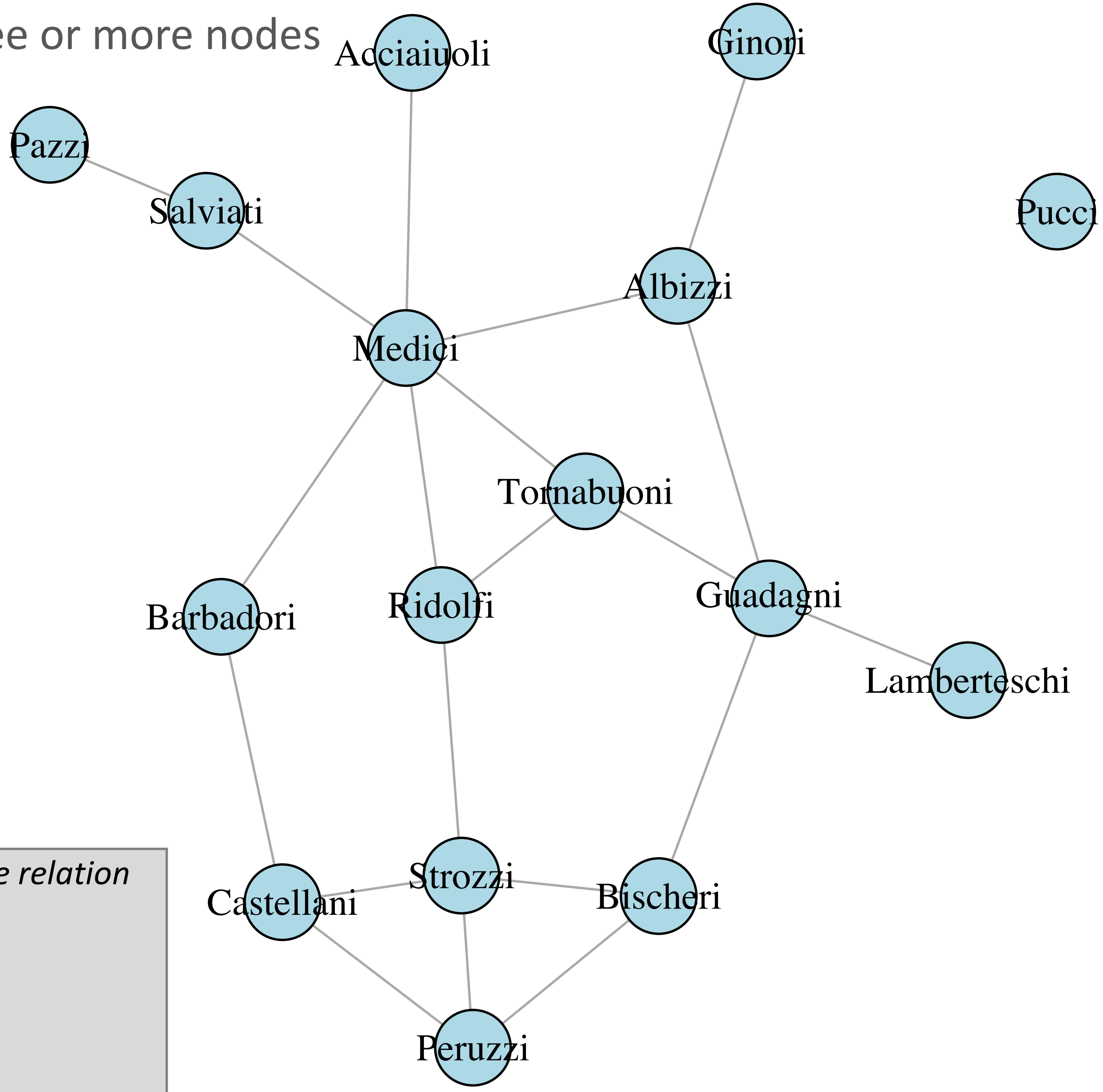


Subgroups Based on Complete Density

Define a **clique** as a **maximal complete** subgraph of three or more nodes

- Cliques may overlap—the same node may belong to more than one clique
- No clique can be entirely contained within another clique (otherwise the smaller clique would not be maximal)

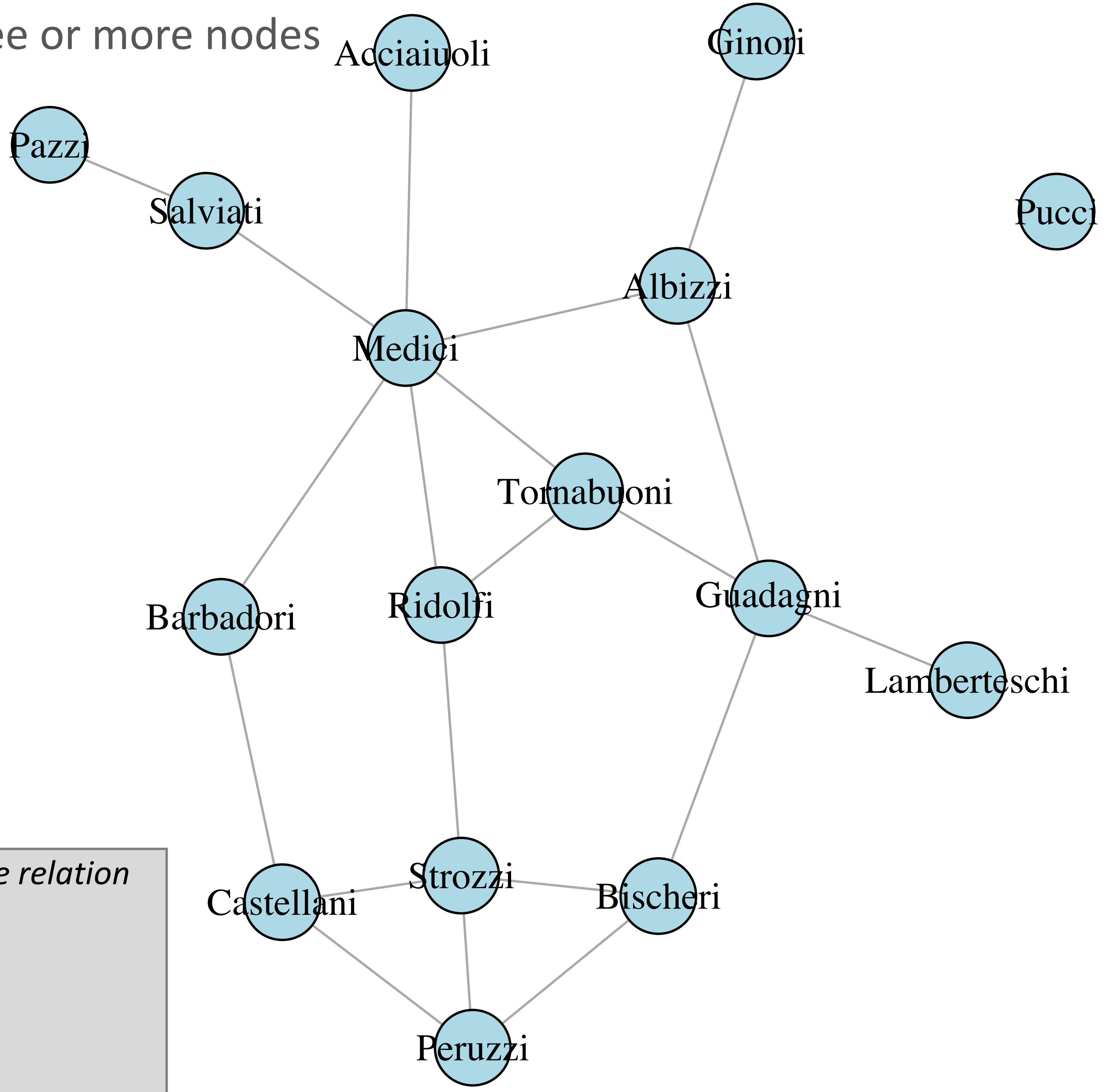
Take 5 minutes to discuss: what are the cliques in the Florentine marriage relation network?



Subgroups Based on Complete Density

Define a **clique** as a **maximal complete** subgraph of three or more nodes

- Cliques may overlap—the same node may belong to more than one clique
- No clique can be entirely contained within another clique (otherwise the smaller clique would not be maximal)

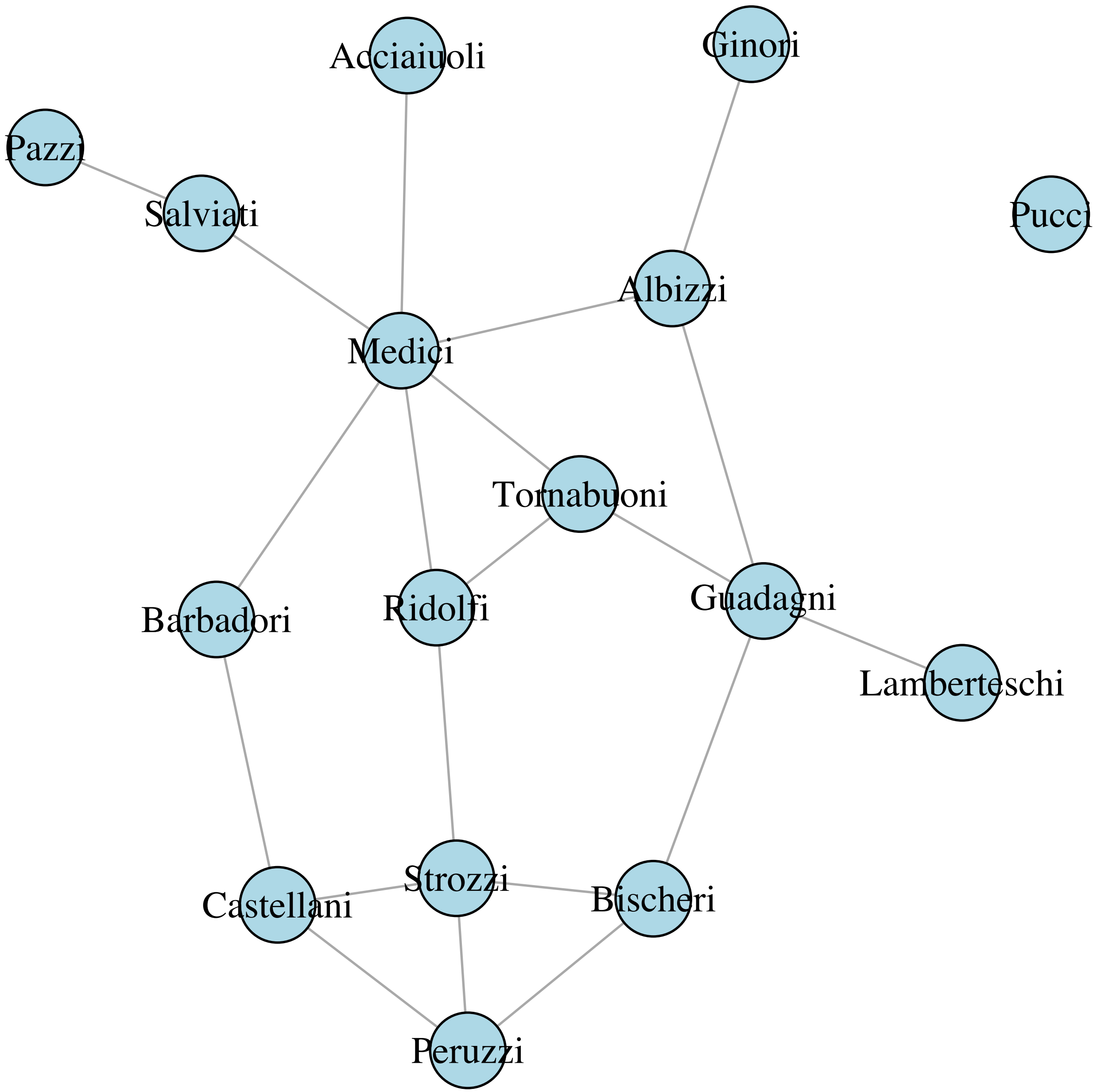


Take 5 minutes to discuss: what are the cliques in the Florentine marriage relation network?
Can identify 3:
Bischeri, Peruzzi, Strozzi
Castellani, Peruzzi, Strozzi
Medici, Ridolfi, Tornabuoni

Subgroups Based on Complete Density

Problems with cliques

- Cliques are "stingy"
- The absence of a single relationship invalidates the entire clique
- A sparse network (like marriage) will have very few cliques
- The size of cliques is limited by the degree of the nodes– if actors are limited to k ties, there can be no clique in the data that has more than $k + 1$ members



Relaxing the Requirement of Complete Density – Reachability/Distance

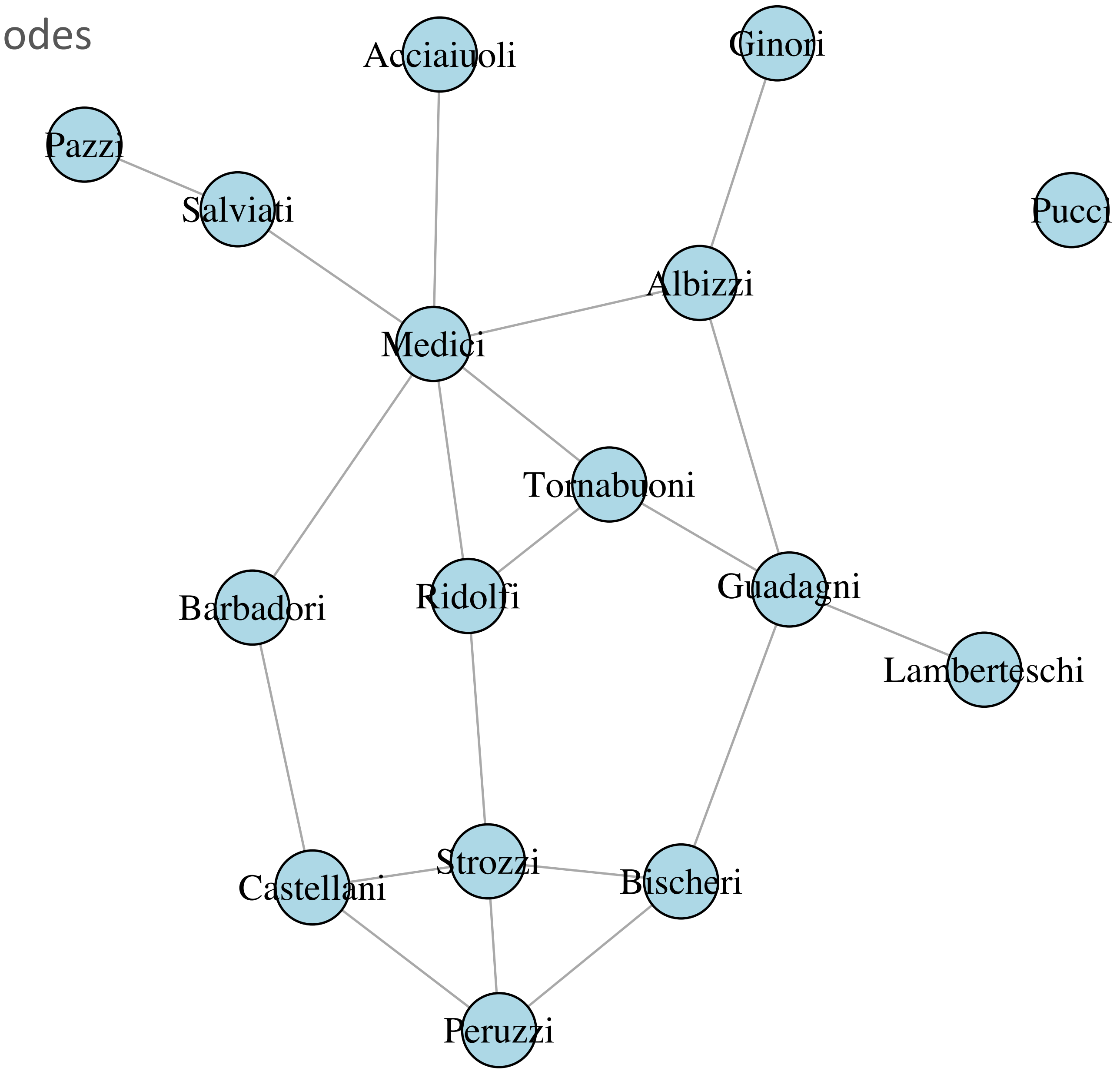
Define an ***n-clique*** as a **maximal** subgraph where no nodes have a shortest path greater than ***n***

- Formally, we can use the notion of geodesic distance to define the nodes $\mathcal{V}_s$ in an n-*clique*:

$$d(i,j) \leq n$$

for all $v_i, v_j \in \mathcal{V}_s$

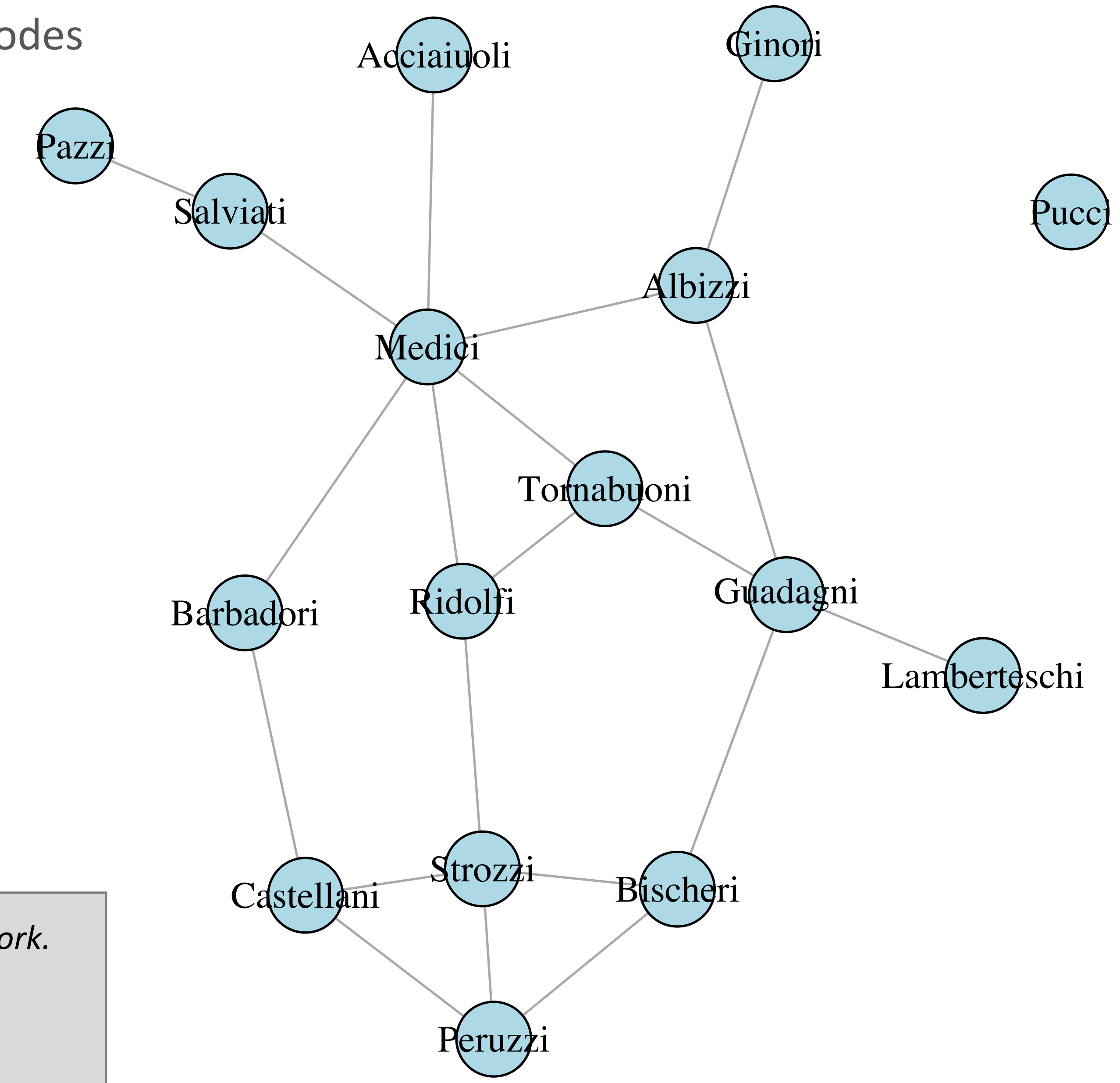
- And there are no additional nodes that are also distance n or less from all the nodes in the subgraph



Relaxing the Requirement of Complete Density – Reachability/Distance

Define an ***n-clique*** as a **maximal** subgraph where no nodes have a shortest path greater than $\boldsymbol{n}$

- When $n = 1$, the subgraph is a clique
- A cutoff value of $n = 2$ is often very useful—in real life, this means that everyone in the group need not know each other, but should be able to reach each other through at least one intermediary



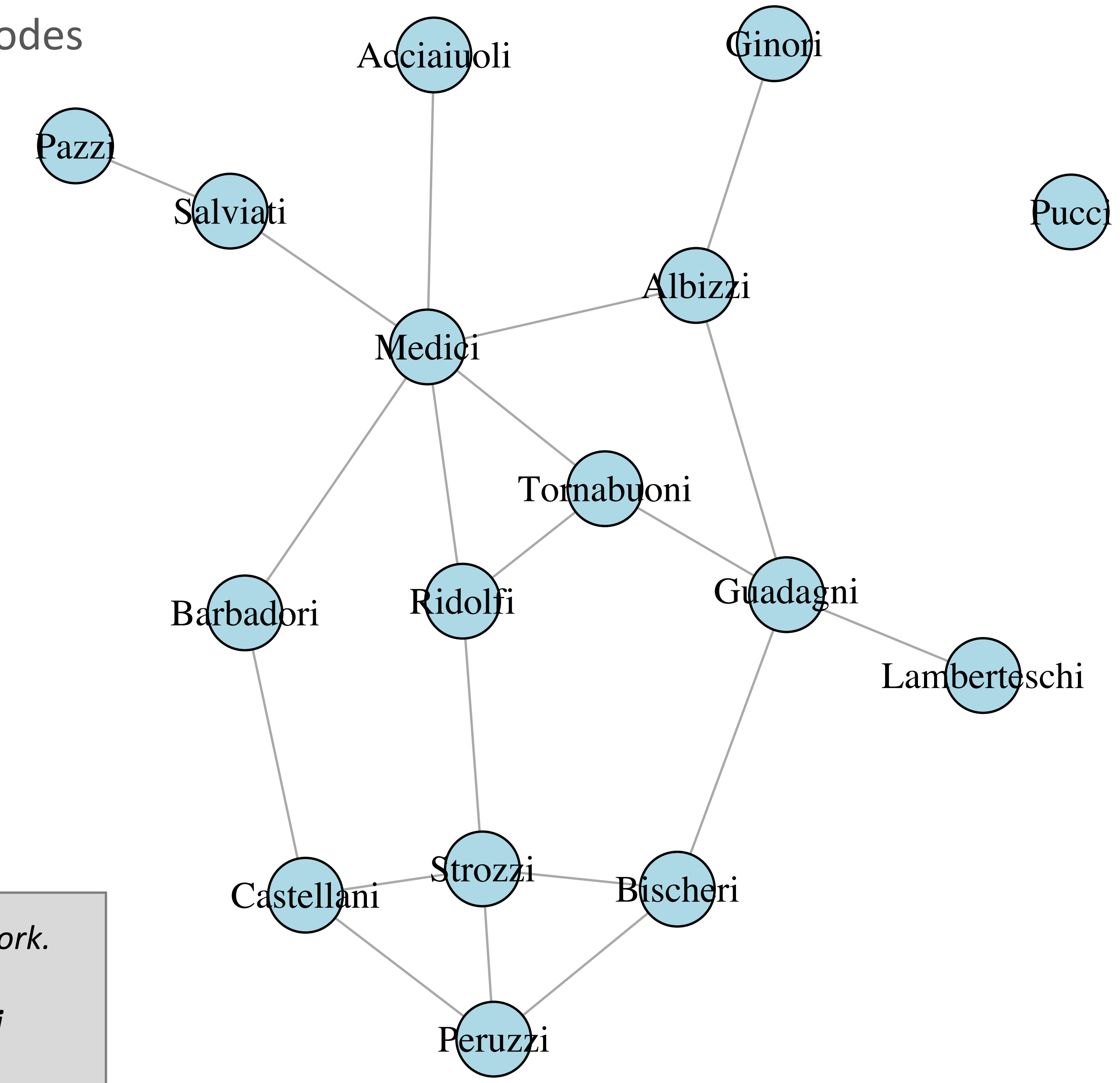
Take 5 minutes to discuss: there are 13 2-cliques in the marriage network. Which one is the largest? Which is the smallest?

Relaxing the Requirement of Complete Density – Reachability/Distance

Define an ***n*-clique** as a **maximal** subgraph where no nodes have a shortest path greater than n

- When $n = 1$, the subgraph is a clique
- A cutoff value of $n = 2$ is often very useful—in real life, this means that everyone in the group need not know each other, but should be able to reach each other through at least one intermediary

Take 5 minutes to discuss: there are 13 2-cliques in the marriage network. Which one is the largest? Which is the smallest?
Acciaiuoli, Albizzi, Barbadori, Medici, Ridolfi, Salviati, Tornabuoni
Medici, Pazzi, Salviati



Relaxing the Requirement of Complete Density – Familiarity/Degree

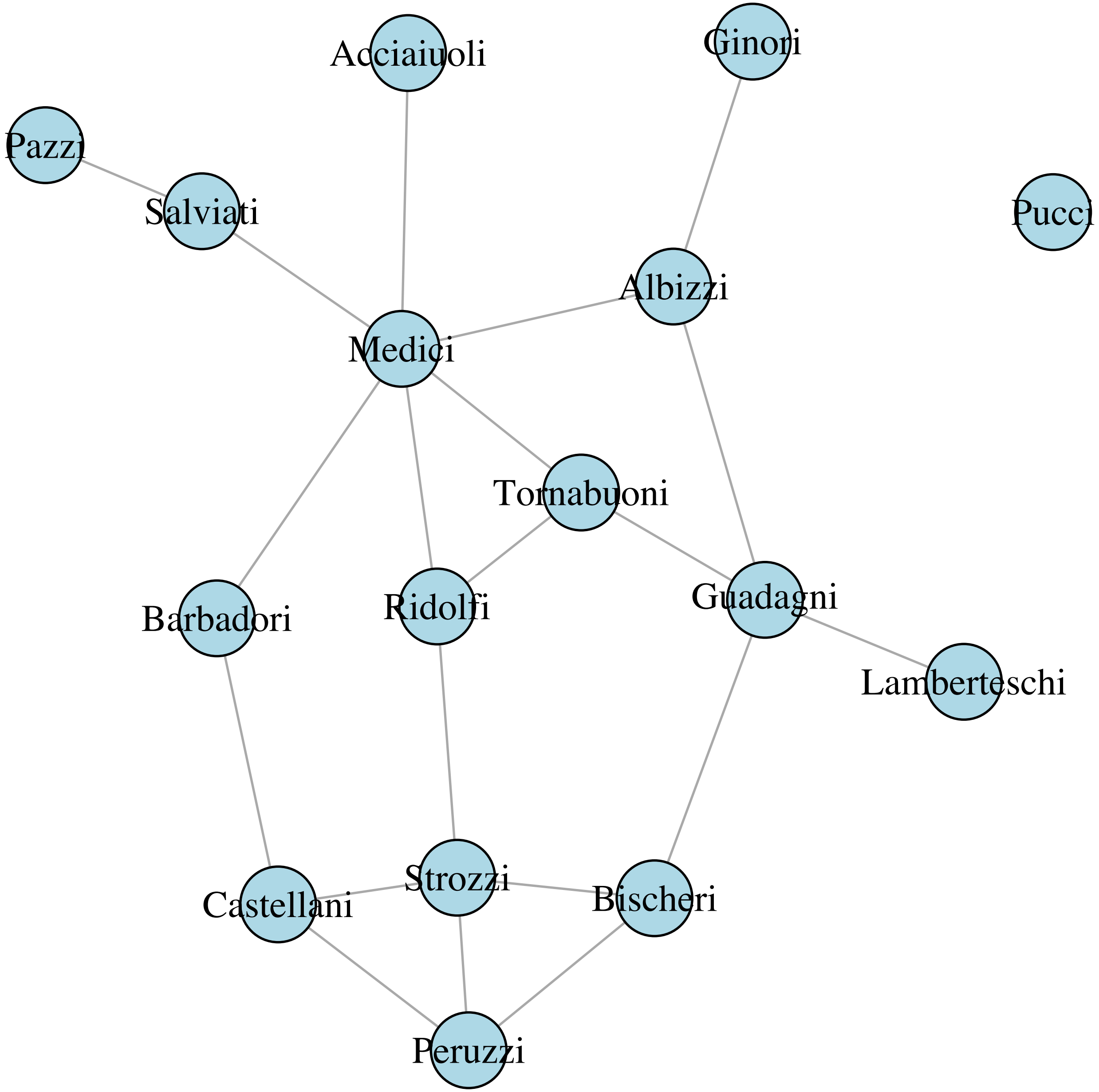
Define a ***k-plex*** as a **maximal** subgraph containing $\boldsymbol{g_s}$ nodes that connect to no fewer than $\boldsymbol{g_s - k}$ nodes in the subgraph

- We can rephrase this to say that each node in the subgraph can be lacking ties to no more than k subgraph members
- Formally, we can use degree to define the nodes $\mathcal{V}_s$ in a k-*plex:*

$$d_s(i) \geq (g_s - k)$$

for all $v_i \in \mathcal{V}_s$

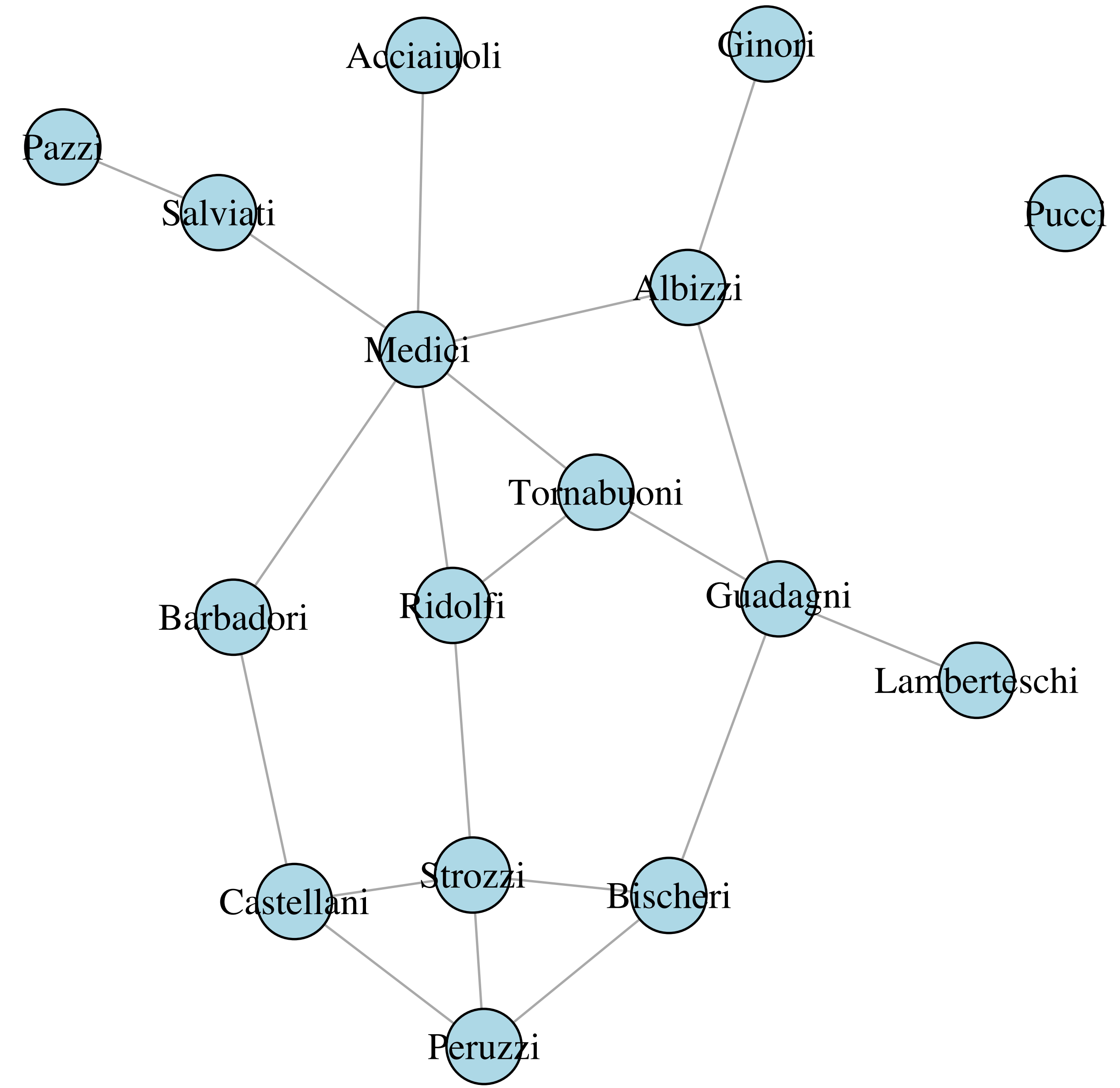
- And there are no additional nodes for which the condition holds



Relaxing the Requirement of Complete Density – Familiarity/Degree

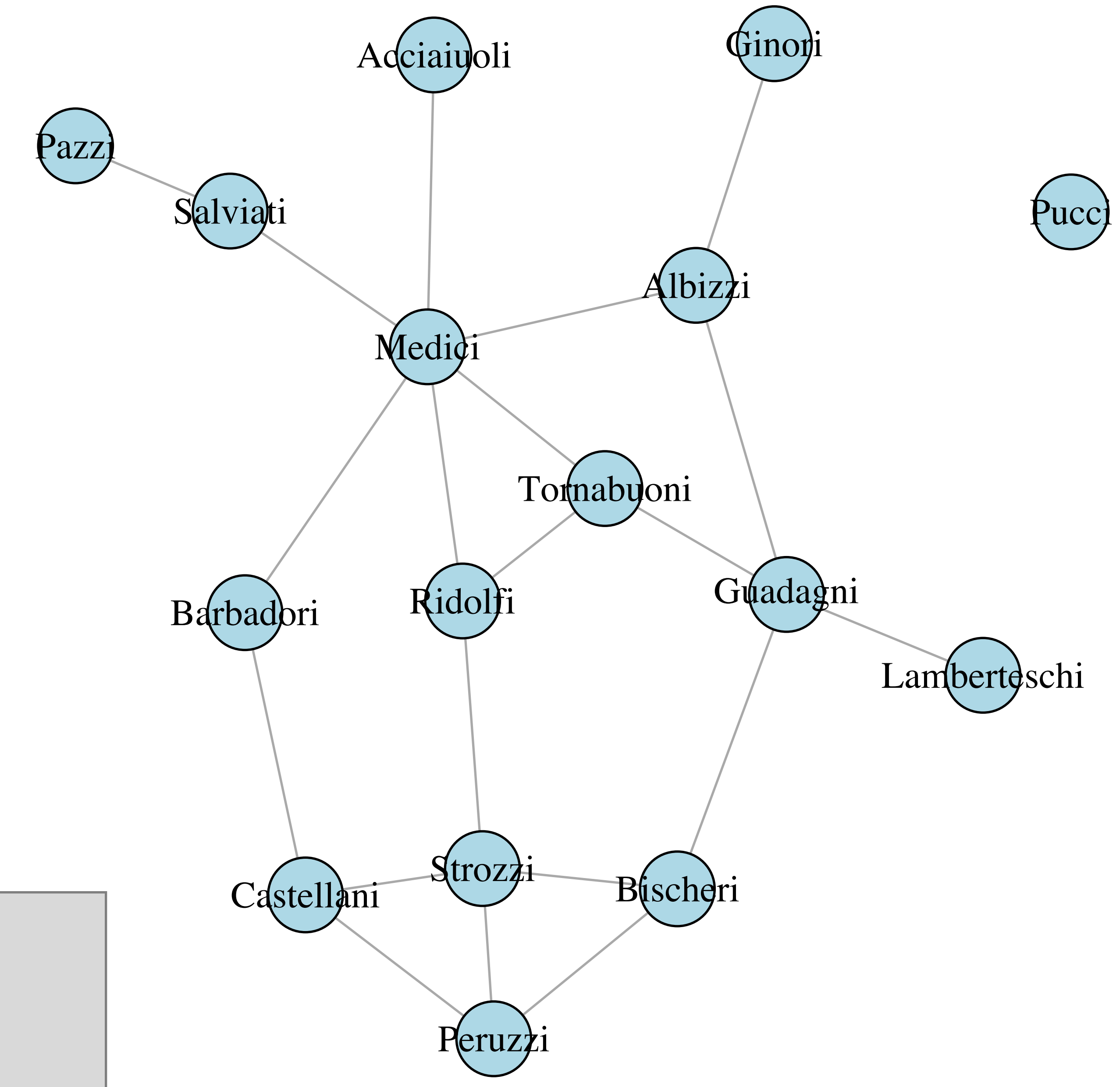
Define a ***k-plex*** as a **maximal** subgraph containing $\boldsymbol{g_s}$ nodes that connect to no fewer than $\boldsymbol{g_s - k}$ nodes in the subgraph

- When $k = 1$, the subgraph is a clique
- As k gets larger, each node is allowed more missing lines in the subgraph
- k-*plexes* are generally more stable than cliques because the removal of one link does not destroy the group
- If you take any subset of k nodes in a k-*plex*, and then consider the k nodes next to them, this is a full census of all nodes in the k-*plex*



Relaxing the Requirement of Complete Density – Familiarity/Degree

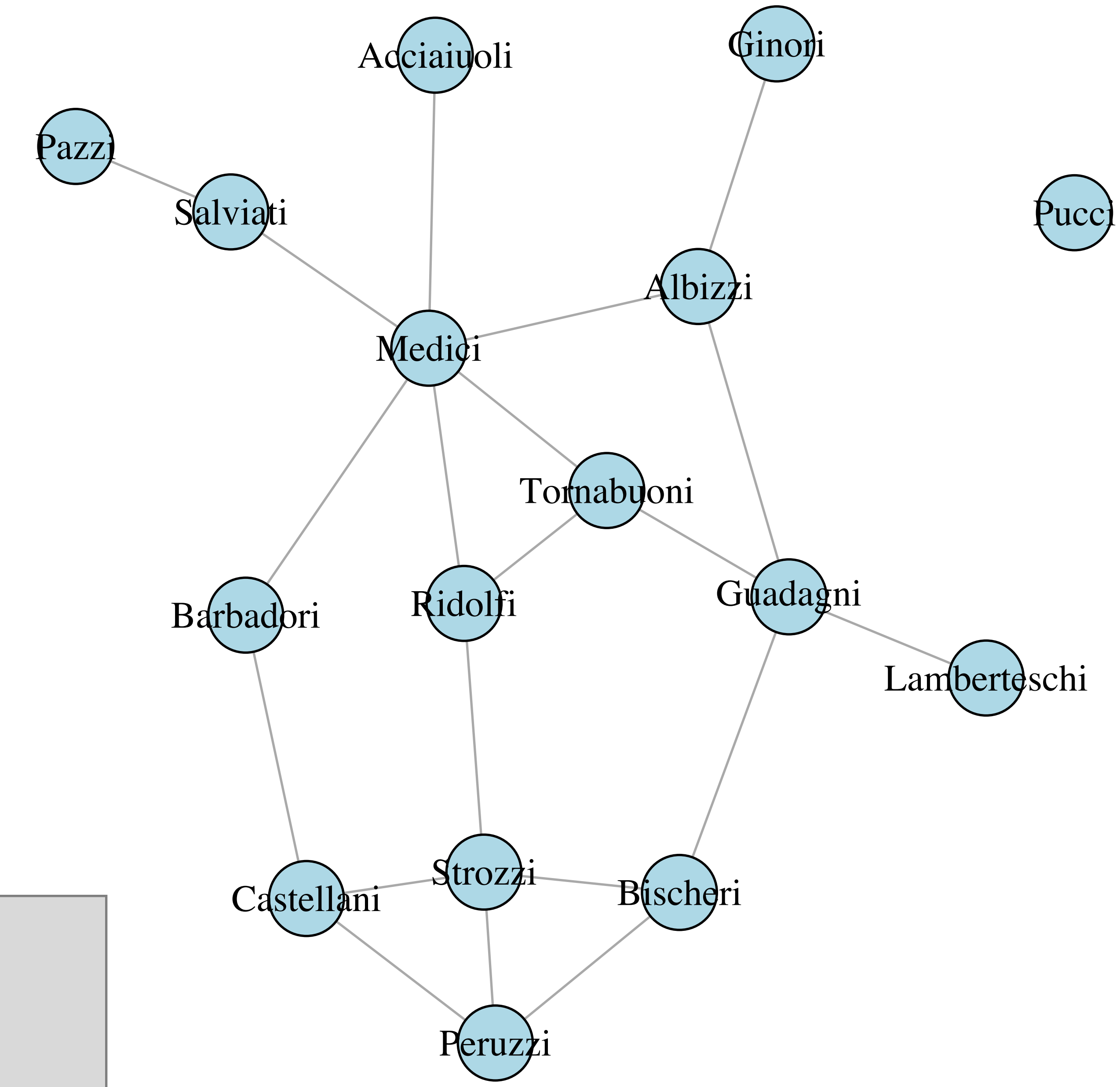
Define a ***k-plex*** as a **maximal** subgraph containing $\boldsymbol{g_s}$ nodes that connect to no fewer than $\boldsymbol{g_s - k}$ nodes in the subgraph



Take 5 minutes to discuss: find a 2-plex in the marriage network.

Relaxing the Requirement of Complete Density – Familiarity/Degree

Define a ***k-plex*** as a **maximal** subgraph containing $\boldsymbol{g_s}$ nodes that connect to no fewer than $\boldsymbol{g_s - k}$ nodes in the subgraph



Take 5 minutes to discuss: find a 2-plex in the marriage network.
Albizzi, Guadagni, Medici, Tornabuoni
Bischeri, Castellani, Peruzzi, Strozzi

Relaxing the Requirement of Complete Density – Familiarity/Degree

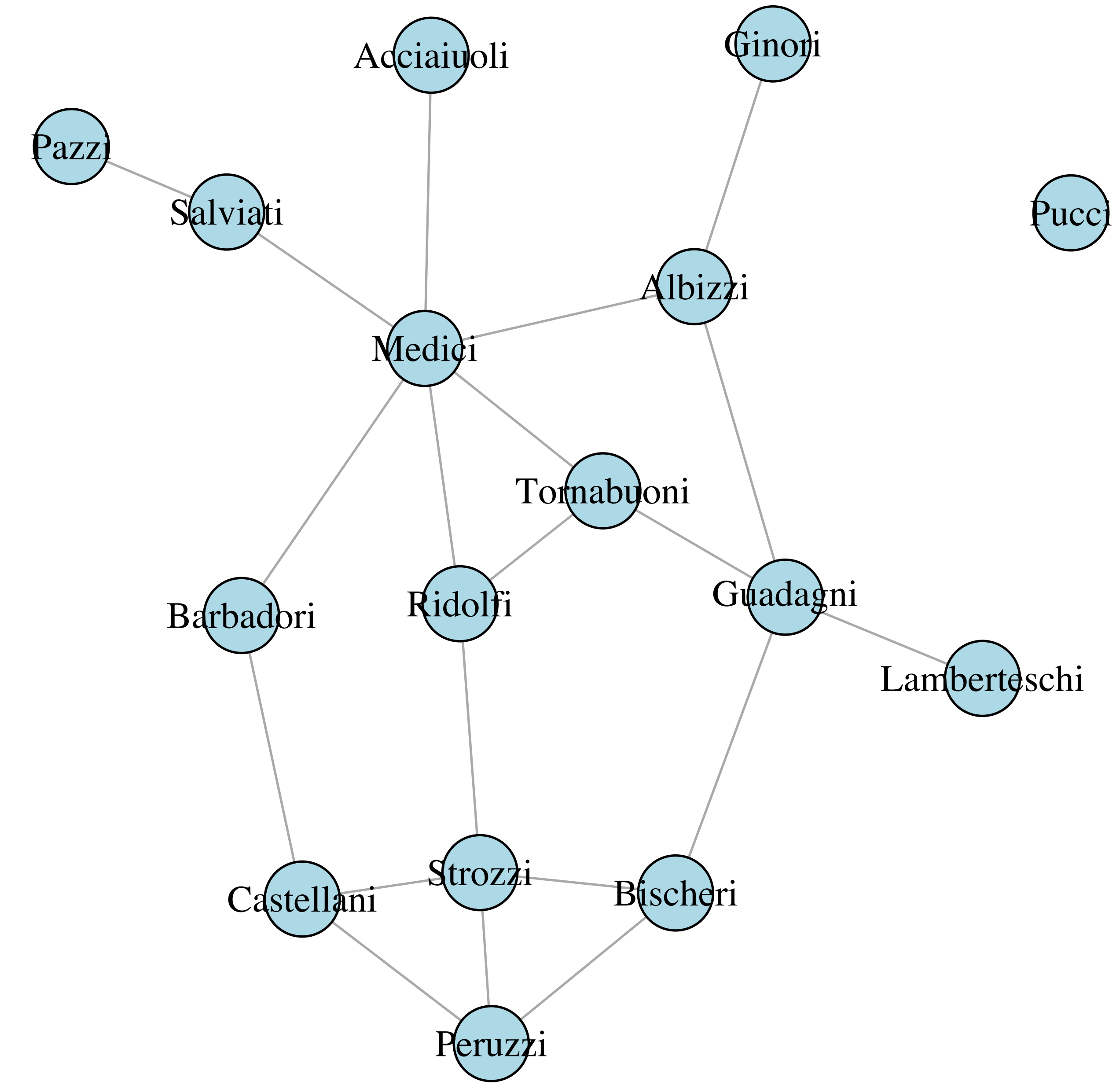
Define a ***k*-core** as a **maximal** subgraph containing $\boldsymbol{g_s}$ nodes that connect to at least $\boldsymbol{k}$ nodes in the subgraph

- As opposed to the k-*plex*, which specifies the number of relationships that can be *absent*, the k-*core* specifies the minimum number that are *present*
- Formally, we can use degree to define the nodes $\mathcal{V}_s$ in a k-*core:*

$$d_s(i) \geq k$$

$$\text{for all } v_i \in \mathcal{V}_s$$

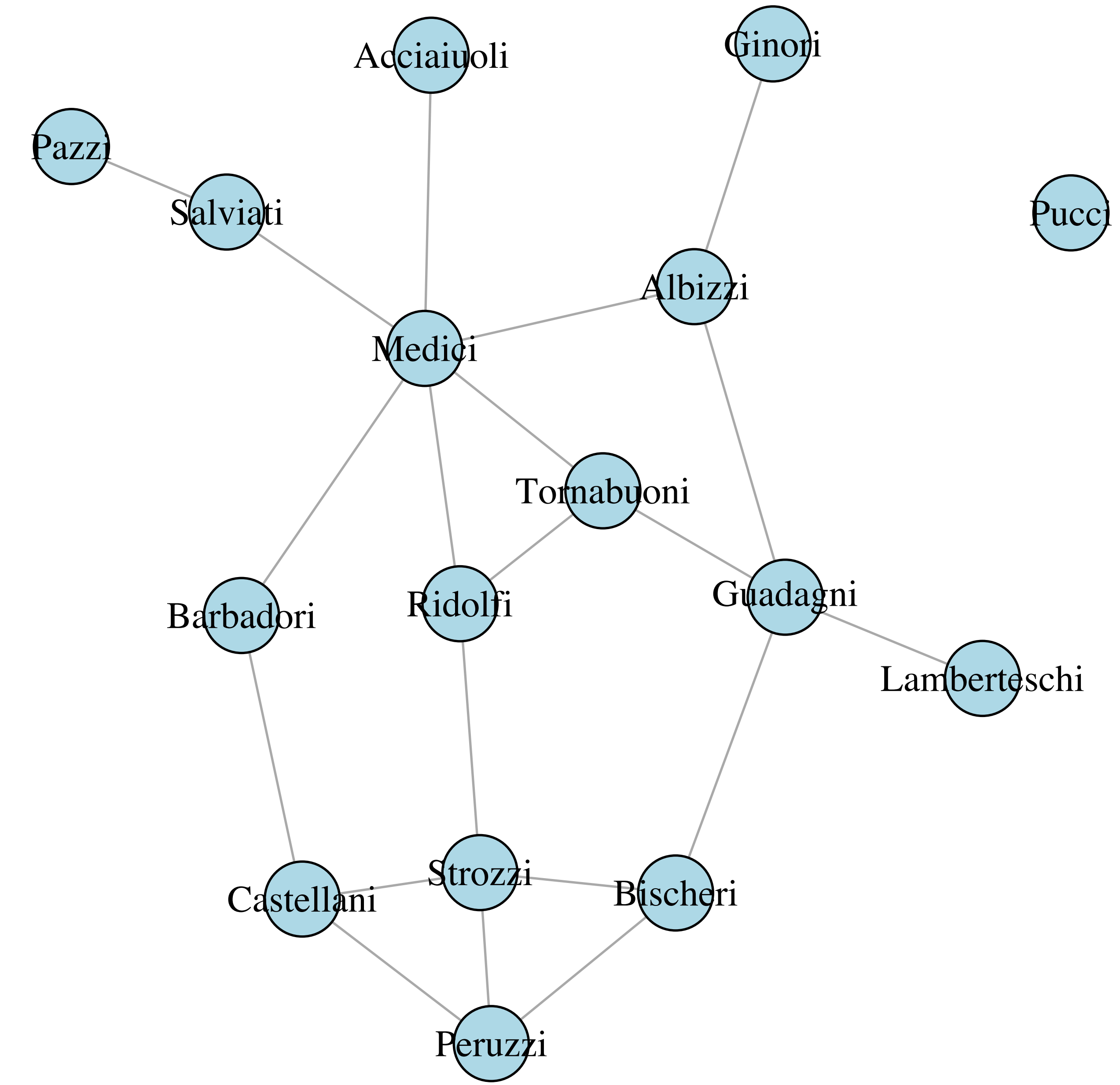
- And there are no additional nodes for which the condition holds



Relaxing the Requirement of Complete Density – Robustness/Connectivity

Define *lambda sets as nodes whose line connectivity in the set is greater than connectivity with any node outside it*

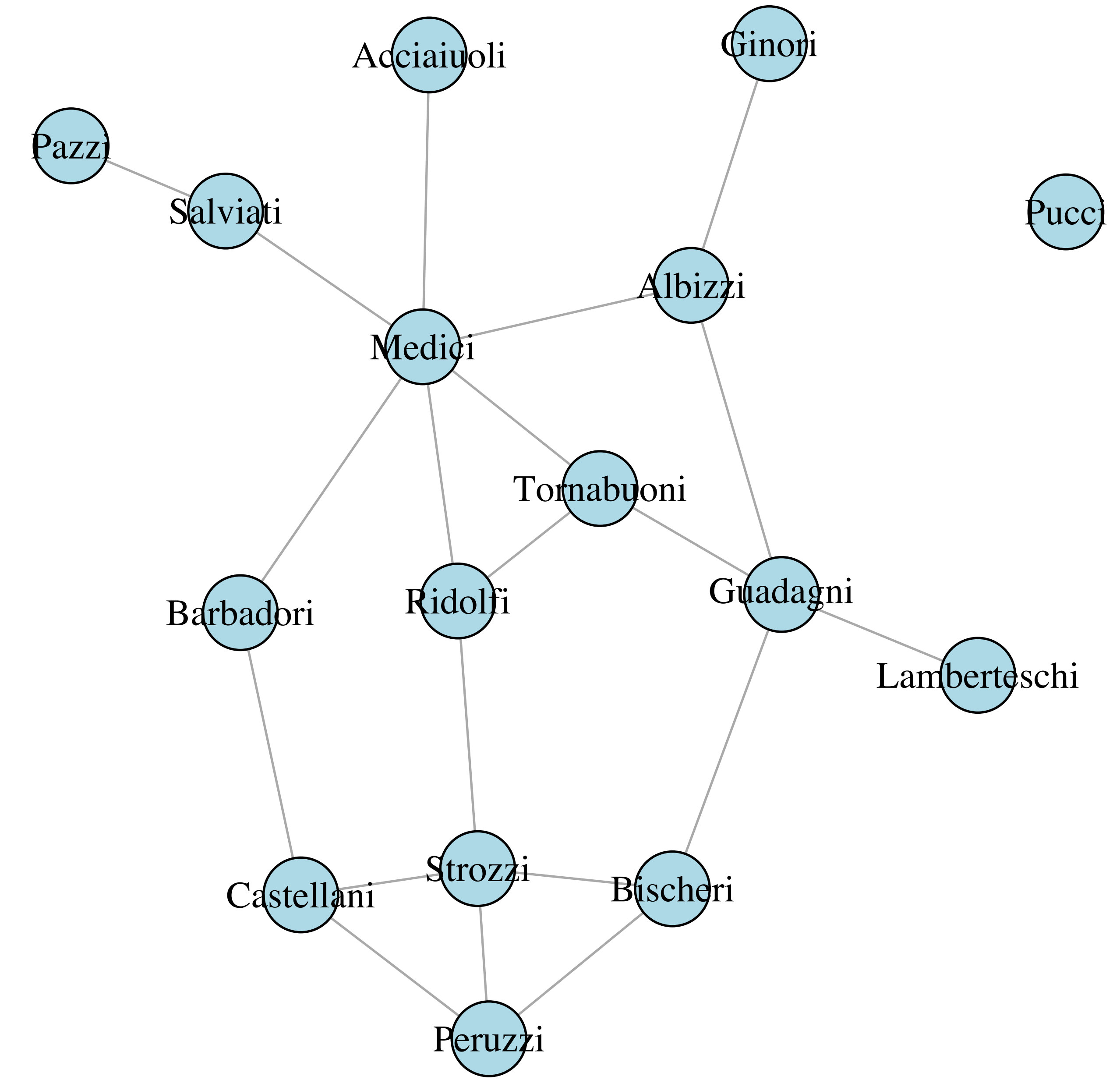
- We define **line connectivity**, $\lambda(i, j)$, as the number of lines that must be removed from the graph in order to leave no path between i and j
- This value is also equal to the the number of paths between two nodes that do not share any of the same lines—these are also called **line-independent paths**



Relaxing the Requirement of Complete Density – Robustness/Connectivity

Define *lambda sets as nodes whose line connectivity in the set is greater than connectivity with any node outside it*

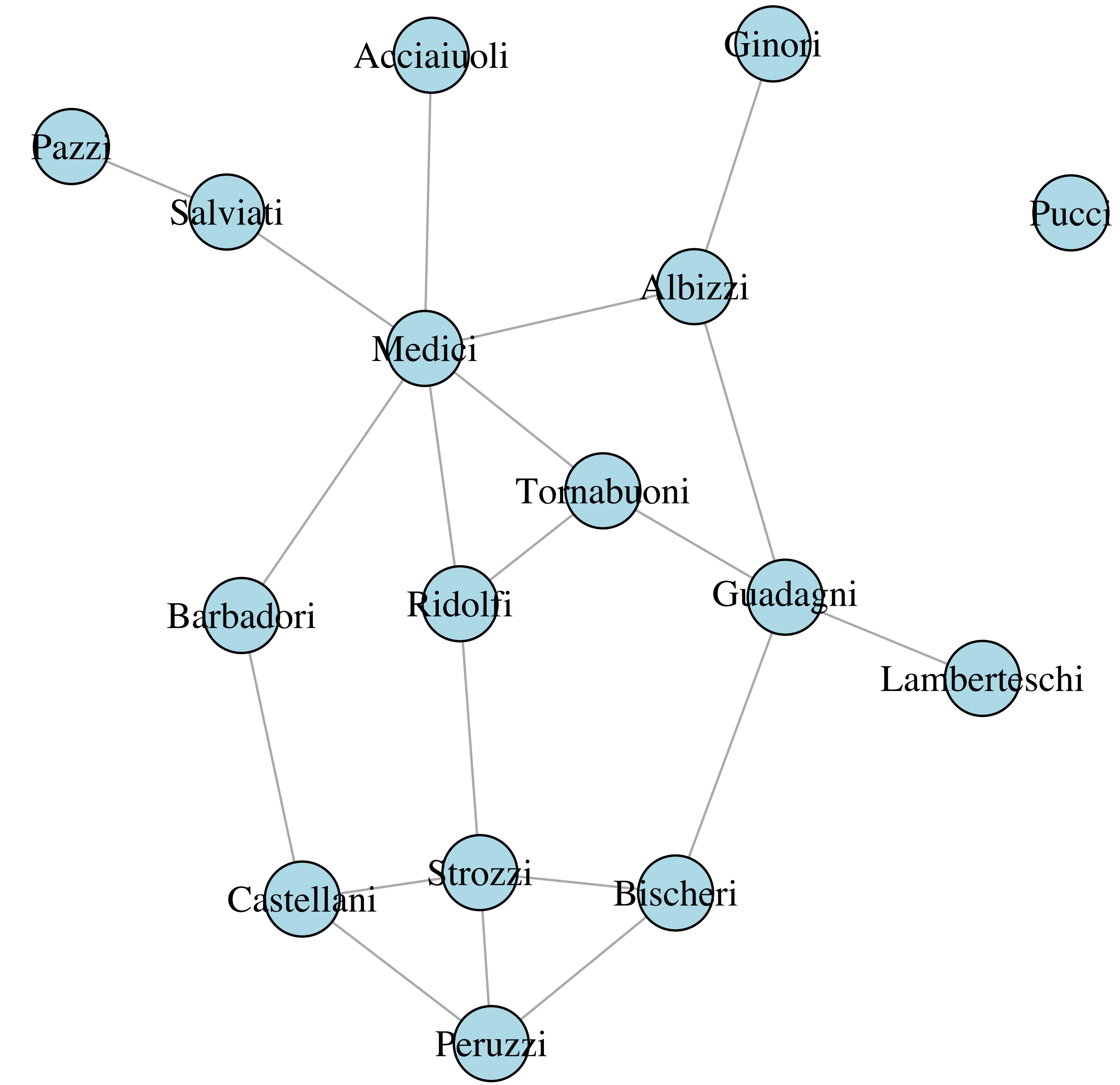
- A set is a *lambda set* if any pair of nodes in the set has a higher connectivity than any pair of nodes consisting of one not within the set and one node outside of the set
- Formally, this is represented as a subset of nodes $\mathcal{V}_s \in \mathcal{V}$ and $i, j, k \in \mathcal{V}_s$, and $l \in \mathcal{V} - \mathcal{V}_s$, $\lambda(i, j) > \lambda(k, l)$



Relaxing the Requirement of Complete Density – Robustness/Connectivity

Define *lambda sets as nodes whose line connectivity in the set is greater than connectivity with any node outside it*

- Important to note that *lambda sets* are not necessarily cohesive in adjacency or distance
- Because there is no restriction on the length of paths that connect nodes in a *lambda set*, members may actually be very distant in the graph



Measuring Tendency Towards Cliques and Defining Clusters

We Can Use the Path Distances Between Each Family to Identify Clustering Tendencies

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	0	2	2	4	3	3	3	4	1	3	4	2	2	3	2
Albizzi	2	0	2	2	3	1	1	2	1	3	3	2	2	3	2
Barbadori	2	2	0	3	1	3	3	4	1	3	2	2	2	2	2
Bischeri	4	2	3	0	2	3	1	2	3	5	1	2	4	1	2
Castellani	3	3	1	2	0	4	3	4	2	4	1	2	3	1	3
Ginori	3	1	3	3	4	0	2	3	2	4	4	3	3	4	3
Guadagni	3	1	3	1	3	2	0	1	2	4	2	2	3	2	1
Lambertes chi	4	2	4	2	4	3	1	0	3	5	3	3	4	3	2
Medici	1	1	1	3	2	2	2	3	0	2	3	1	1	2	1
Pazzi	3	3	3	5	4	4	4	5	2	0	5	3	1	4	3
Peruzzi	4	3	2	1	1	4	2	3	3	5	0	2	4	1	3
Ridolfi	2	2	2	2	2	3	2	3	1	3	2	0	2	1	1
Salviati	2	2	2	4	3	3	3	4	1	1	4	2	0	3	2
Strozzi	3	3	2	1	1	4	2	3	2	4	1	1	3	0	2
Tornabuon i	2	2	2	2	3	3	1	2	1	3	3	1	2	2	0

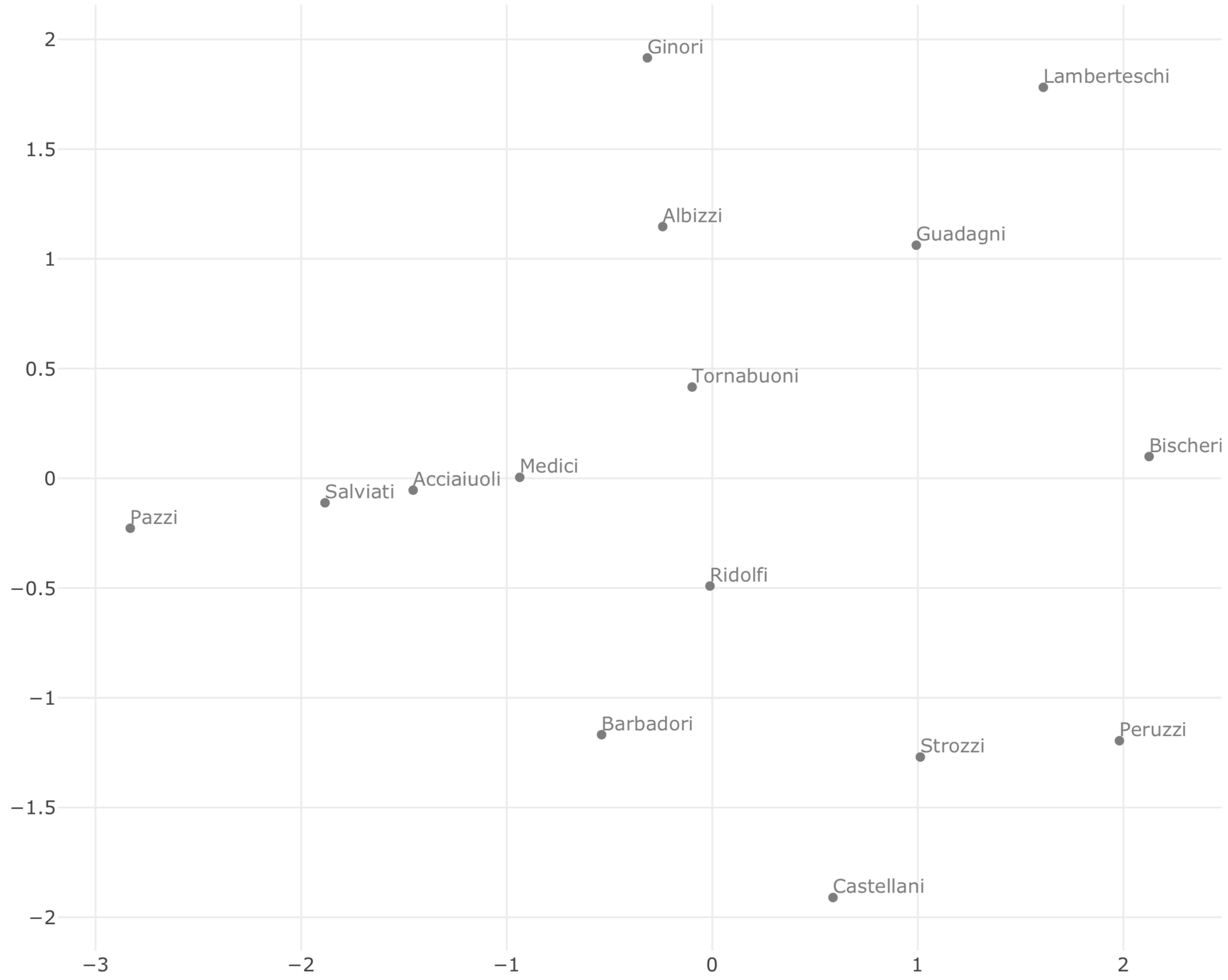
Using Path Distance to View Clustering Tendency

We can use multidimensional scaling to view the members of a network in lower, e.g., 2-dimensional space

- Multidimensional scaling helps to show similarities or dissimilarities so that more similar nodes are closer to each other, and more dissimilar nodes are far away
- The scaling shows which nodes are relatively close to each other in a graph theoretic sense

Using Path Distance to View Clustering Tendency

We can use multidimensional scaling to view the members of a network in lower, e.g., 2-dimensional space



- Shows that Medici family is near the center
- Also places the 6 families in the historically anti-Medici faction (Bischeri, Castellani, Cuadagni, Lamberteschi, Peruzzi, and Strozzi)

  to the right of the plot, away from the Medicis

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

Start with the original matrix, and then take a correlation among the rows or columns

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Albizzi	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0
Barbadori	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
Bischeri	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0
Castellani	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0
Ginori	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Guadagni	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1
Lambertes chi	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Medici	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1
Pazzi	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Peruzzi	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0
Ridolfi	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
Salviati	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
Strozzi	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0
Tornabuon i	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

Then take another correlation

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.535	0.681	-0.134	-0.134	-0.071	-0.161	-0.071	-0.218	-0.071	-0.134	0.535	0.681	-0.161	0.535
Albizzi	0.535	1	0.294	0.167	-0.25	-0.134	-0.302	0.535	-0.408	-0.134	-0.25	0.167	0.294	-0.302	0.583
Barbadori	0.681	0.294	1	-0.196	-0.196	-0.105	-0.237	-0.105	-0.32	-0.105	0.294	0.294	0.423	0.207	0.294
Bischeri	-0.134	0.167	-0.196	1	0.583	-0.134	-0.302	0.535	-0.408	-0.134	0.167	0.167	-0.196	0.075	0.167
Castellani	-0.134	-0.25	-0.196	0.583	1	-0.134	-0.302	-0.134	-0.068	-0.134	0.167	0.167	-0.196	0.075	-0.25
Ginori	-0.071	-0.134	-0.105	-0.134	-0.134	1	0.443	-0.071	0.327	-0.071	-0.134	-0.134	-0.105	-0.161	-0.134
Guadagni	-0.161	-0.302	-0.237	-0.302	-0.302	0.443	1	-0.161	0.123	-0.161	0.075	0.075	-0.237	-0.023	-0.302
Lambertes chi	-0.071	0.535	-0.105	0.535	-0.134	-0.071	-0.161	1	-0.218	-0.071	-0.134	-0.134	-0.105	-0.161	0.535
Medici	-0.218	-0.408	-0.32	-0.408	-0.068	0.327	0.123	-0.218	1	0.327	-0.408	-0.068	-0.32	-0.185	-0.068
Pazzi	-0.071	-0.134	-0.105	-0.134	-0.134	-0.071	-0.161	-0.071	0.327	1	-0.134	-0.134	-0.105	-0.161	-0.134
Peruzzi	-0.134	-0.25	0.294	0.167	0.167	-0.134	0.075	-0.134	-0.408	-0.134	1	0.167	-0.196	0.452	-0.25
Ridolfi	0.535	0.167	0.294	0.167	0.167	-0.134	0.075	-0.134	-0.068	-0.134	0.167	1	0.294	-0.302	0.167
Salviati	0.681	0.294	0.423	-0.196	-0.196	-0.105	-0.237	-0.105	-0.32	-0.105	-0.196	0.294	1	-0.237	0.294
Strozzi	-0.161	-0.302	0.207	0.075	0.075	-0.161	-0.023	-0.161	-0.185	-0.161	0.452	-0.302	-0.237	1	0.075
Tornabuon i	0.535	0.583	0.294	0.167	-0.25	-0.134	-0.302	0.535	-0.068	-0.134	-0.25	0.167	0.294	0.075	1

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

And another

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.718	0.83	-0.233	-0.372	-0.332	-0.387	0.026	-0.404	-0.233	-0.251	0.669	0.931	-0.332	0.673
Albizzi	0.718	1	0.496	0.216	-0.343	-0.379	-0.529	0.676	-0.54	-0.259	-0.38	0.336	0.635	-0.404	0.875
Barbadori	0.83	0.496	1	-0.249	-0.326	-0.408	-0.368	-0.122	-0.546	-0.287	0.214	0.484	0.748	0.133	0.48
Bischeri	-0.233	0.216	-0.249	1	0.692	-0.424	-0.464	0.611	-0.512	-0.303	0.242	0.041	-0.253	0.127	0.152
Castellani	-0.372	-0.343	-0.326	0.692	1	-0.265	-0.288	-0.088	-0.147	-0.156	0.378	0.109	-0.364	0.226	-0.383
Ginori	-0.332	-0.379	-0.408	-0.424	-0.265	1	0.727	-0.249	0.598	0.046	-0.275	-0.318	-0.299	-0.239	-0.401
Guadagni	-0.387	-0.529	-0.368	-0.464	-0.288	0.727	1	-0.394	0.417	-0.068	0.089	-0.133	-0.38	-0.009	-0.557
Lambertes chi	0.026	0.676	-0.122	0.611	-0.088	-0.249	-0.394	1	-0.361	-0.171	-0.284	-0.157	-0.012	-0.206	0.651
Medici	-0.404	-0.54	-0.546	-0.512	-0.147	0.598	0.417	-0.361	1	0.566	-0.468	-0.31	-0.414	-0.257	-0.371
Pazzi	-0.233	-0.259	-0.287	-0.303	-0.156	0.046	-0.068	-0.171	0.566	1	-0.303	-0.304	-0.201	-0.239	-0.249
Peruzzi	-0.251	-0.38	0.214	0.242	0.378	-0.275	0.089	-0.284	-0.468	-0.303	1	0.07	-0.274	0.705	-0.422
Ridolfi	0.669	0.336	0.484	0.041	0.109	-0.318	-0.133	-0.157	-0.31	-0.304	0.07	1	0.573	-0.387	0.226
Salviati	0.931	0.635	0.748	-0.253	-0.364	-0.299	-0.38	-0.012	-0.414	-0.201	-0.274	0.573	1	-0.357	0.578
Strozzi	-0.332	-0.404	0.133	0.127	0.226	-0.239	-0.009	-0.206	-0.257	-0.239	0.705	-0.387	-0.357	1	-0.192
Tornabuon i	0.673	0.875	0.48	0.152	-0.383	-0.401	-0.557	0.651	-0.371	-0.249	-0.422	0.226	0.578	-0.192	1

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

And another

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.866	0.919	-0.131	-0.528	-0.558	-0.614	0.305	-0.593	-0.412	-0.352	0.845	0.997	-0.476	0.848
Albizzi	0.866	1	0.727	0.231	-0.409	-0.646	-0.778	0.739	-0.684	-0.448	-0.441	0.643	0.852	-0.493	0.986
Barbadori	0.919	0.727	1	-0.091	-0.398	-0.655	-0.593	0.156	-0.722	-0.533	0.006	0.789	0.911	-0.108	0.72
Bischeri	-0.131	0.231	-0.091	1	0.726	-0.605	-0.574	0.644	-0.592	-0.456	0.358	0.03	-0.156	0.299	0.202
Castellani	-0.528	-0.409	-0.398	0.726	1	-0.245	-0.115	-0.034	-0.163	-0.168	0.651	-0.146	-0.538	0.545	-0.442
Ginori	-0.558	-0.646	-0.655	-0.605	-0.245	1	0.919	-0.51	0.879	0.494	-0.249	-0.562	-0.535	-0.158	-0.641
Guadagni	-0.614	-0.778	-0.593	-0.574	-0.115	0.919	1	-0.676	0.761	0.347	0.084	-0.501	-0.599	0.106	-0.783
Lambertes chi	0.305	0.739	0.156	0.644	-0.034	-0.51	-0.676	1	-0.527	-0.328	-0.338	0.103	0.281	-0.269	0.741
Medici	-0.593	-0.684	-0.722	-0.592	-0.163	0.879	0.761	-0.527	1	0.807	-0.343	-0.605	-0.572	-0.201	-0.649
Pazzi	-0.412	-0.448	-0.533	-0.456	-0.168	0.494	0.347	-0.328	0.807	1	-0.397	-0.52	-0.387	-0.251	-0.416
Peruzzi	-0.352	-0.441	0.006	0.358	0.651	-0.249	0.084	-0.338	-0.343	-0.397	1	-0.068	-0.367	0.894	-0.459
Ridolfi	0.845	0.643	0.789	0.03	-0.146	-0.562	-0.501	0.103	-0.605	-0.52	-0.068	1	0.836	-0.382	0.574
Salviati	0.997	0.852	0.911	-0.156	-0.538	-0.535	-0.599	0.281	-0.572	-0.387	-0.367	0.836	1	-0.492	0.831
Strozzi	-0.476	-0.493	-0.108	0.299	0.545	-0.158	0.106	-0.269	-0.201	-0.251	0.894	-0.382	-0.492	1	-0.444
Tornabuon i	0.848	0.986	0.72	0.202	-0.442	-0.641	-0.783	0.741	-0.649	-0.416	-0.459	0.574	0.831	-0.444	1

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

And another

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.961	0.97	0.112	-0.621	-0.76	-0.831	0.693	-0.767	-0.67	-0.489	0.965	1	-0.643	0.958
Albizzi	0.961	1	0.919	0.316	-0.506	-0.833	-0.92	0.866	-0.829	-0.713	-0.478	0.914	0.956	-0.61	0.999
Barbadori	0.97	0.919	1	0.211	-0.478	-0.837	-0.852	0.648	-0.856	-0.787	-0.27	0.974	0.966	-0.442	0.916
Bischeri	0.112	0.316	0.211	1	0.638	-0.701	-0.624	0.641	-0.68	-0.666	0.467	0.226	0.091	0.394	0.306
Castellani	-0.621	-0.506	-0.478	0.638	1	-0.013	0.15	-0.168	0.004	-0.082	0.864	-0.445	-0.636	0.867	-0.518
Ginori	-0.76	-0.833	-0.837	-0.701	-0.013	1	0.971	-0.814	0.985	0.902	-0.059	-0.822	-0.746	0.1	-0.827
Guadagni	-0.831	-0.92	-0.852	-0.624	0.15	0.971	1	-0.904	0.942	0.819	0.169	-0.845	-0.82	0.309	-0.917
Lambertes chi	0.693	0.866	0.648	0.641	-0.168	-0.814	-0.904	1	-0.789	-0.669	-0.328	0.644	0.682	-0.4	0.868
Medici	-0.767	-0.829	-0.856	-0.68	0.004	0.985	0.942	-0.789	1	0.959	-0.094	-0.837	-0.753	0.075	-0.821
Pazzi	-0.67	-0.713	-0.787	-0.666	-0.082	0.902	0.819	-0.669	0.959	1	-0.238	-0.775	-0.655	-0.056	-0.701
Peruzzi	-0.489	-0.478	-0.27	0.467	0.864	-0.059	0.169	-0.328	-0.094	-0.238	1	-0.298	-0.503	0.967	-0.489
Ridolfi	0.965	0.914	0.974	0.226	-0.445	-0.822	-0.845	0.644	-0.837	-0.775	-0.298	1	0.962	-0.498	0.903
Salviati	1	0.956	0.966	0.091	-0.636	-0.746	-0.82	0.682	-0.753	-0.655	-0.503	0.962	1	-0.655	0.954
Strozzi	-0.643	-0.61	-0.442	0.394	0.867	0.1	0.309	-0.4	0.075	-0.056	0.967	-0.498	-0.655	1	-0.613
Tornabuon i	0.958	0.999	0.916	0.306	-0.518	-0.827	-0.917	0.868	-0.821	-0.701	-0.489	0.903	0.954	-0.613	1

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

And another

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.994	0.992	0.438	-0.732	-0.905	-0.949	0.941	-0.903	-0.862	-0.667	0.993	1	-0.808	0.994
Albizzi	0.994	1	0.992	0.519	-0.674	-0.936	-0.973	0.972	-0.933	-0.894	-0.621	0.993	0.993	-0.768	1
Barbadori	0.992	0.992	1	0.53	-0.647	-0.946	-0.973	0.954	-0.946	-0.916	-0.567	1	0.99	-0.727	0.991
Bischeri	0.438	0.519	0.53	1	0.28	-0.777	-0.699	0.681	-0.775	-0.815	0.304	0.525	0.426	0.123	0.513
Castellani	-0.732	-0.674	-0.647	0.28	1	0.374	0.487	-0.51	0.37	0.289	0.975	-0.653	-0.742	0.978	-0.679
Ginori	-0.905	-0.936	-0.946	-0.777	0.374	1	0.991	-0.97	0.999	0.992	0.306	-0.944	-0.899	0.494	-0.933
Guadagni	-0.949	-0.973	-0.973	-0.699	0.487	0.991	1	-0.991	0.988	0.967	0.431	-0.972	-0.944	0.605	-0.972
Lambertes chi	0.941	0.972	0.954	0.681	-0.51	-0.97	-0.991	1	-0.965	-0.935	-0.483	0.956	0.937	-0.641	0.971
Medici	-0.903	-0.933	-0.946	-0.775	0.37	0.999	0.988	-0.965	1	0.995	0.297	-0.944	-0.897	0.487	-0.93
Pazzi	-0.862	-0.894	-0.916	-0.815	0.289	0.992	0.967	-0.935	0.995	1	0.205	-0.913	-0.855	0.403	-0.891
Peruzzi	-0.667	-0.621	-0.567	0.304	0.975	0.306	0.431	-0.483	0.297	0.205	1	-0.576	-0.676	0.977	-0.626
Ridolfi	0.993	0.993	1	0.525	-0.653	-0.944	-0.972	0.956	-0.944	-0.913	-0.576	1	0.992	-0.735	0.993
Salviati	1	0.993	0.99	0.426	-0.742	-0.899	-0.944	0.937	-0.897	-0.855	-0.676	0.992	1	-0.816	0.993
Strozzi	-0.808	-0.768	-0.727	0.123	0.978	0.494	0.605	-0.641	0.487	0.403	0.977	-0.735	-0.816	1	-0.773
Tornabuon i	0.994	1	0.991	0.513	-0.679	-0.933	-0.972	0.971	-0.93	-0.891	-0.626	0.993	0.993	-0.773	1

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

And another

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.999	0.998	0.778	-0.905	-0.978	-0.989	0.992	-0.978	-0.967	-0.879	0.999	1	-0.944	0.999
Albizzi	0.999	1	1	0.801	-0.889	-0.985	-0.994	0.996	-0.985	-0.975	-0.861	1	0.999	-0.932	1
Barbadori	0.998	1	1	0.812	-0.88	-0.988	-0.996	0.997	-0.988	-0.979	-0.851	1	0.998	-0.924	1
Bischeri	0.778	0.801	0.812	1	-0.438	-0.892	-0.861	0.851	-0.893	-0.913	-0.386	0.81	0.774	-0.529	0.799
Castellani	-0.905	-0.889	-0.88	-0.438	1	0.797	0.834	-0.845	0.795	0.766	0.998	-0.882	-0.908	0.994	-0.89
Ginori	-0.978	-0.985	-0.988	-0.892	0.797	1	0.998	-0.996	1	0.999	0.76	-0.988	-0.977	0.855	-0.985
Guadagni	-0.989	-0.994	-0.996	-0.861	0.834	0.998	1	-1	0.998	0.993	0.801	-0.996	-0.988	0.887	-0.994
Lambertes chi	0.992	0.996	0.997	0.851	-0.845	-0.996	-1	1	-0.996	-0.991	-0.813	0.997	0.991	-0.896	0.996
Medici	-0.978	-0.985	-0.988	-0.893	0.795	1	0.998	-0.996	1	0.999	0.759	-0.987	-0.976	0.854	-0.984
Pazzi	-0.967	-0.975	-0.979	-0.913	0.766	0.999	0.993	-0.991	0.999	1	0.727	-0.979	-0.965	0.828	-0.975
Peruzzi	-0.879	-0.861	-0.851	-0.386	0.998	0.76	0.801	-0.813	0.759	0.727	1	-0.853	-0.882	0.987	-0.862
Ridolfi	0.999	1	1	0.81	-0.882	-0.988	-0.996	0.997	-0.987	-0.979	-0.853	1	0.998	-0.926	1
Salviati	1	0.999	0.998	0.774	-0.908	-0.977	-0.988	0.991	-0.976	-0.965	-0.882	0.998	1	-0.946	0.999
Strozzi	-0.944	-0.932	-0.924	-0.529	0.994	0.855	0.887	-0.896	0.854	0.828	0.987	-0.926	-0.946	1	-0.933
Tornabuon i	0.999	1	1	0.799	-0.89	-0.985	-0.994	0.996	-0.984	-0.975	-0.862	1	0.999	-0.933	1

We Can Use the Correlations Between Families’ Sets of Relationships to Determine Clustering Tendencies

And another

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	1	1	0.979	-0.992	-0.998	-0.999	0.999	-0.998	-0.998	-0.989	1	1	-0.996	1
Albizzi	1	1	1	0.981	-0.991	-0.999	-1	1	-0.999	-0.998	-0.988	1	1	-0.995	1
Barbadori	1	1	1	0.982	-0.99	-0.999	-1	1	-0.999	-0.998	-0.987	1	1	-0.994	1
Bischeri	0.979	0.981	0.982	1	-0.945	-0.989	-0.986	0.985	-0.989	-0.991	-0.938	0.981	0.978	-0.955	0.98
Castellani	-0.992	-0.991	-0.99	-0.945	1	0.983	0.986	-0.987	0.983	0.981	1	-0.99	-0.992	0.999	-0.991
Ginori	-0.998	-0.999	-0.999	-0.989	0.983	1	1	-1	1	1	0.98	-0.999	-0.998	0.989	-0.999
Guadagni	-0.999	-1	-1	-0.986	0.986	1	1	-1	1	1	0.983	-1	-0.999	0.991	-1
Lambertes chi	0.999	1	1	0.985	-0.987	-1	-1	1	-1	-0.999	-0.984	1	0.999	-0.992	1
Medici	-0.998	-0.999	-0.999	-0.989	0.983	1	1	-1	1	1	0.98	-0.999	-0.998	0.989	-0.999
Pazzi	-0.998	-0.998	-0.998	-0.991	0.981	1	1	-0.999	1	1	0.977	-0.998	-0.997	0.987	-0.998
Peruzzi	-0.989	-0.988	-0.987	-0.938	1	0.98	0.983	-0.984	0.98	0.977	1	-0.987	-0.99	0.999	-0.988
Ridolfi	1	1	1	0.981	-0.99	-0.999	-1	1	-0.999	-0.998	-0.987	1	1	-0.994	1
Salviati	1	1	1	0.978	-0.992	-0.998	-0.999	0.999	-0.998	-0.997	-0.99	1	1	-0.996	1
Strozzi	-0.996	-0.995	-0.994	-0.955	0.999	0.989	0.991	-0.992	0.989	0.987	0.999	-0.994	-0.996	1	-0.995
Tornabuon i	1	1	1	0.98	-0.991	-0.999	-1	1	-0.999	-0.998	-0.988	1	1	-0.995	1

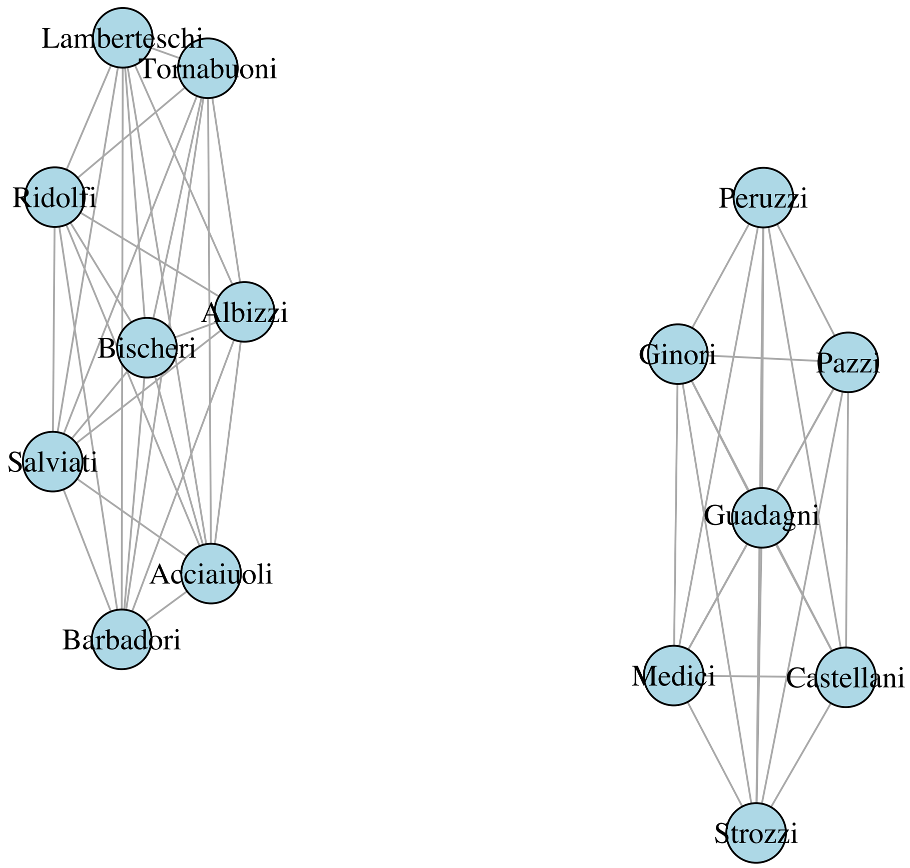
We Can Use the Correlations Between Families’ Sets of Relationships to Determine Clustering Tendencies

After the 8th iteration, we get a matrix of -1s and 1s. This is equivalent to two clusters

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Albizzi	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Barbadori	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Bischeri	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Castellani	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Ginori	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Guadagni	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Lambertes chi	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Medici	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Pazzi	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Peruzzi	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Ridolfi	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Salviati	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1
Strozzi	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	1	-1
Tornabuon i	1	1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1

We Can Use the Correlations Between Families' Sets of Relationships to Determine Clustering Tendencies

After the 8th iteration, we get a matrix of -1s and 1s. This is equivalent to two clusters

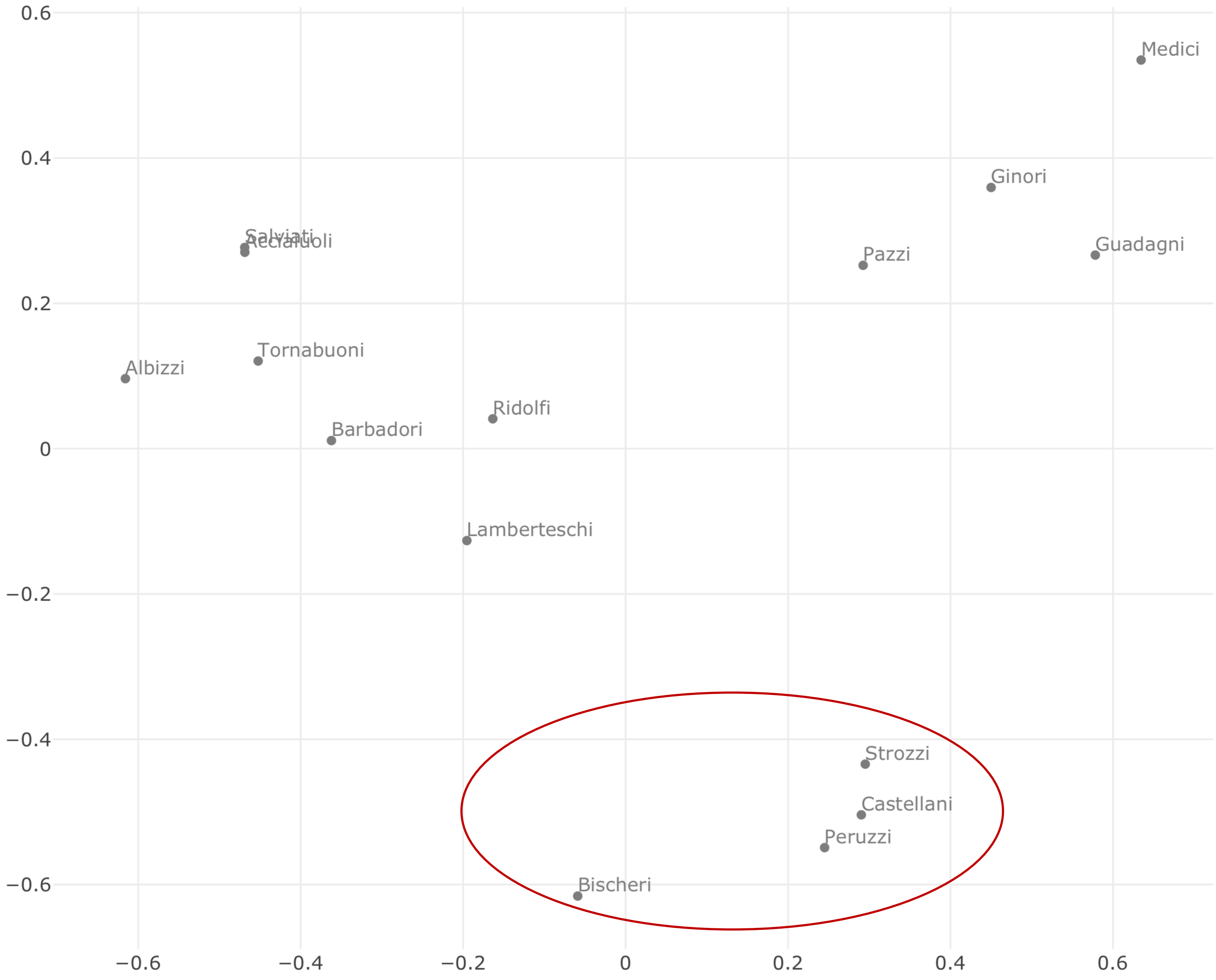


- This time, the anti-Medici families end up inside the same cluster as the Medicis
- Correlational clustering more about equivalent relationship patterns than distance in the network

What if We Run Multidimensional Scaling Using the First Correlation Matrix as the Input, Instead of Distance?

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori	Guadagni	Lambertes chi	Medici	Pazzi	Peruzzi	Ridolfi	Salviati	Strozzi	Tornabuon i
Acciaiuoli	1	0.535	0.681	-0.134	-0.134	-0.071	-0.161	-0.071	-0.218	-0.071	-0.134	0.535	0.681	-0.161	0.535
Albizzi	0.535	1	0.294	0.167	-0.25	-0.134	-0.302	0.535	-0.408	-0.134	-0.25	0.167	0.294	-0.302	0.583
Barbadori	0.681	0.294	1	-0.196	-0.196	-0.105	-0.237	-0.105	-0.32	-0.105	0.294	0.294	0.423	0.207	0.294
Bischeri	-0.134	0.167	-0.196	1	0.583	-0.134	-0.302	0.535	-0.408	-0.134	0.167	0.167	-0.196	0.075	0.167
Castellani	-0.134	-0.25	-0.196	0.583	1	-0.134	-0.302	-0.134	-0.068	-0.134	0.167	0.167	-0.196	0.075	-0.25
Ginori	-0.071	-0.134	-0.105	-0.134	-0.134	1	0.443	-0.071	0.327	-0.071	-0.134	-0.134	-0.105	-0.161	-0.134
Guadagni	-0.161	-0.302	-0.237	-0.302	-0.302	0.443	1	-0.161	0.123	-0.161	0.075	0.075	-0.237	-0.023	-0.302
Lambertes chi	-0.071	0.535	-0.105	0.535	-0.134	-0.071	-0.161	1	-0.218	-0.071	-0.134	-0.134	-0.105	-0.161	0.535
Medici	-0.218	-0.408	-0.32	-0.408	-0.068	0.327	0.123	-0.218	1	0.327	-0.408	-0.068	-0.32	-0.185	-0.068
Pazzi	-0.071	-0.134	-0.105	-0.134	-0.134	-0.071	-0.161	-0.071	0.327	1	-0.134	-0.134	-0.105	-0.161	-0.134
Peruzzi	-0.134	-0.25	0.294	0.167	0.167	-0.134	0.075	-0.134	-0.408	-0.134	1	0.167	-0.196	0.452	-0.25
Ridolfi	0.535	0.167	0.294	0.167	0.167	-0.134	0.075	-0.134	-0.068	-0.134	0.167	1	0.294	-0.302	0.167
Salviati	0.681	0.294	0.423	-0.196	-0.196	-0.105	-0.237	-0.105	-0.32	-0.105	-0.196	0.294	1	-0.237	0.294
Strozzi	-0.161	-0.302	0.207	0.075	0.075	-0.161	-0.023	-0.161	-0.185	-0.161	0.452	-0.302	-0.237	1	0.075
Tornabuon i	0.535	0.583	0.294	0.167	-0.25	-0.134	-0.302	0.535	-0.068	-0.134	-0.25	0.167	0.294	0.075	1

What if We Run Multidimensional Scaling Using the First Correlation Matrix as the Input, Instead of Distance?



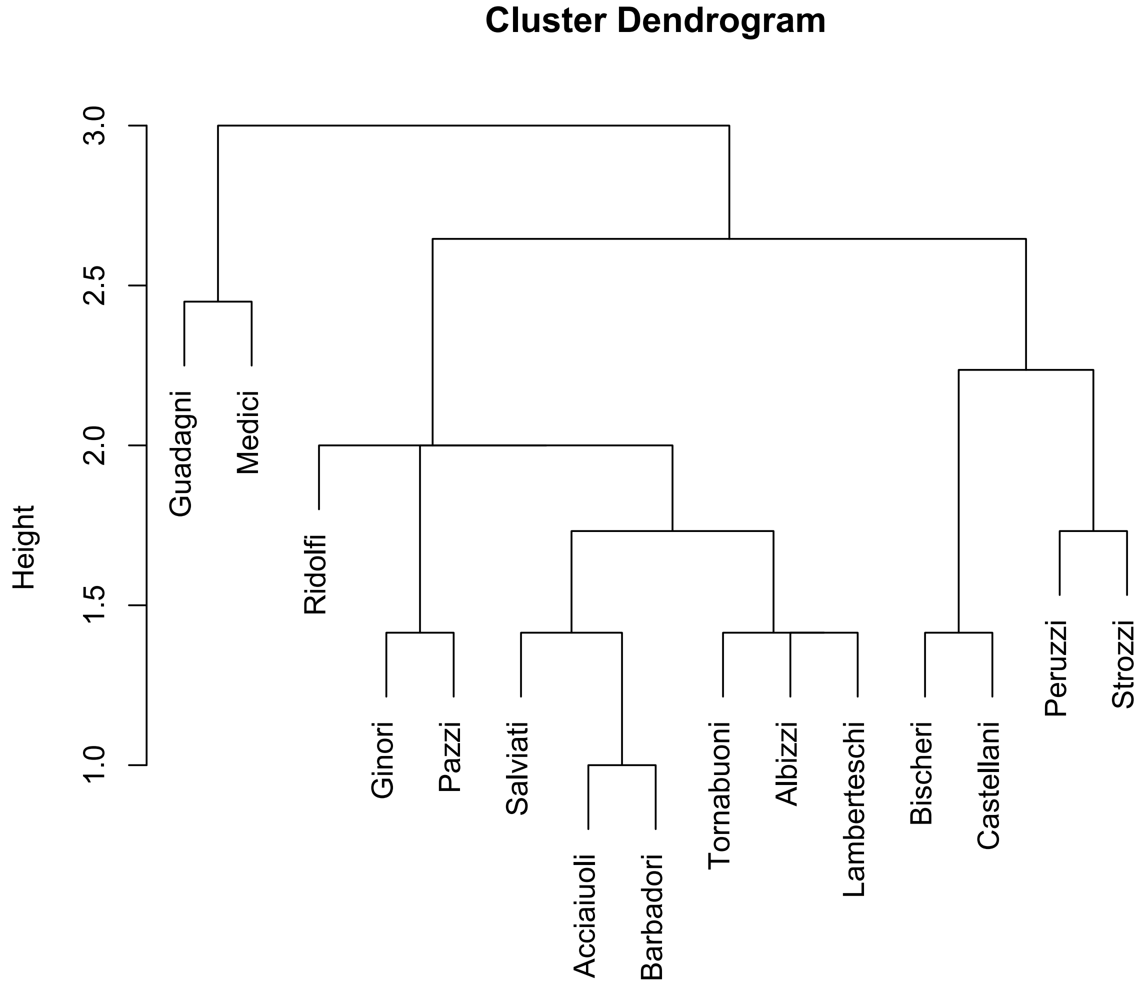
- The anti-Medici faction is in a cluster together, not too close, but not too far, from the Medicis

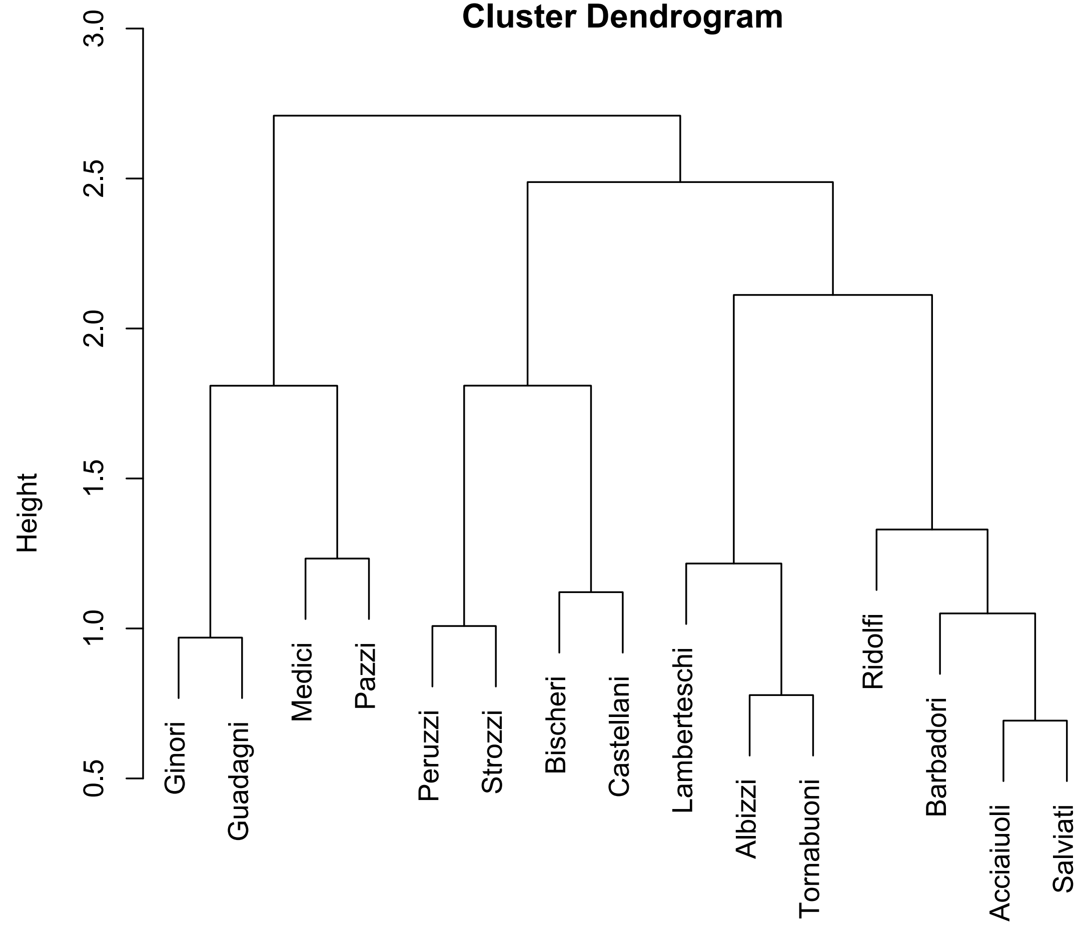
Hierarchical, k-*means*/medoid, and Model-based Clustering

Hierarchical Clustering Can Be Based on Distance or Correlation

- Want to define more specific subsets of nodes that are more specific within each successive cluster
- E.g., decide on a threshold value of α such that pairs of nodes within a cluster satisfy $d_{ij} \leq \alpha$, for the distance measure, or $r_{ij} \leq \alpha$, for the correlation measure
- The procedure then iterates through the desired number of steps of α to produce a clustering tree, or dendogram

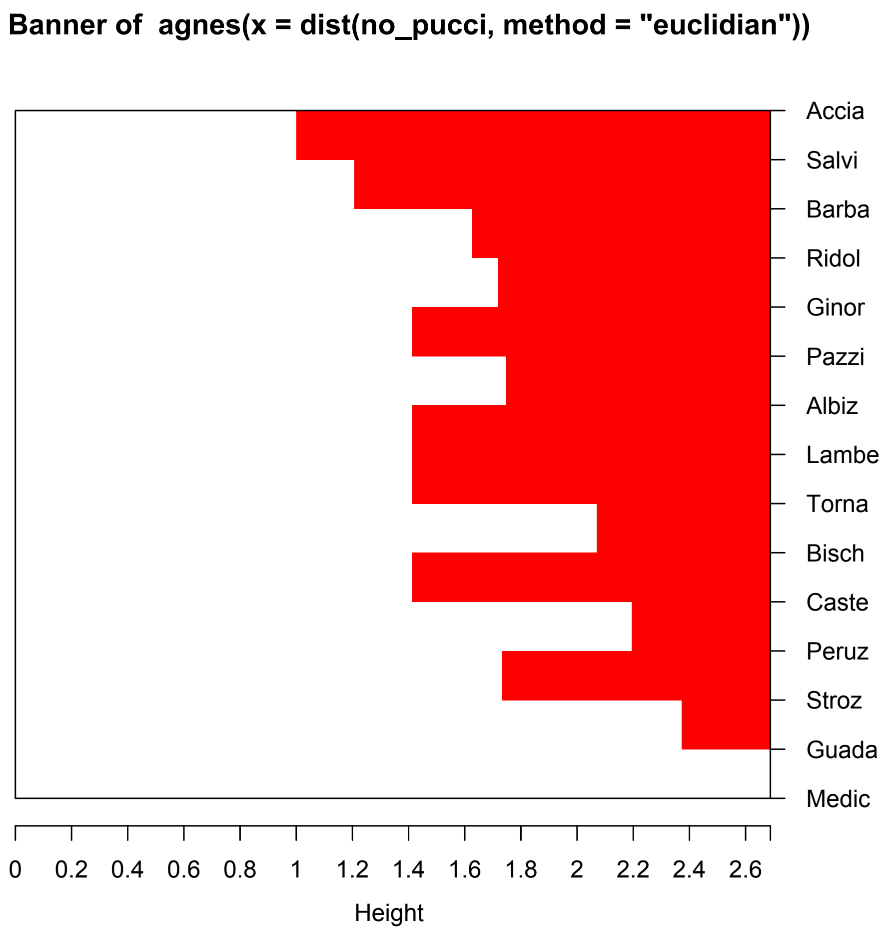
Hierarchical Clustering Can Be Based on Distance (left) or Correlation (right)

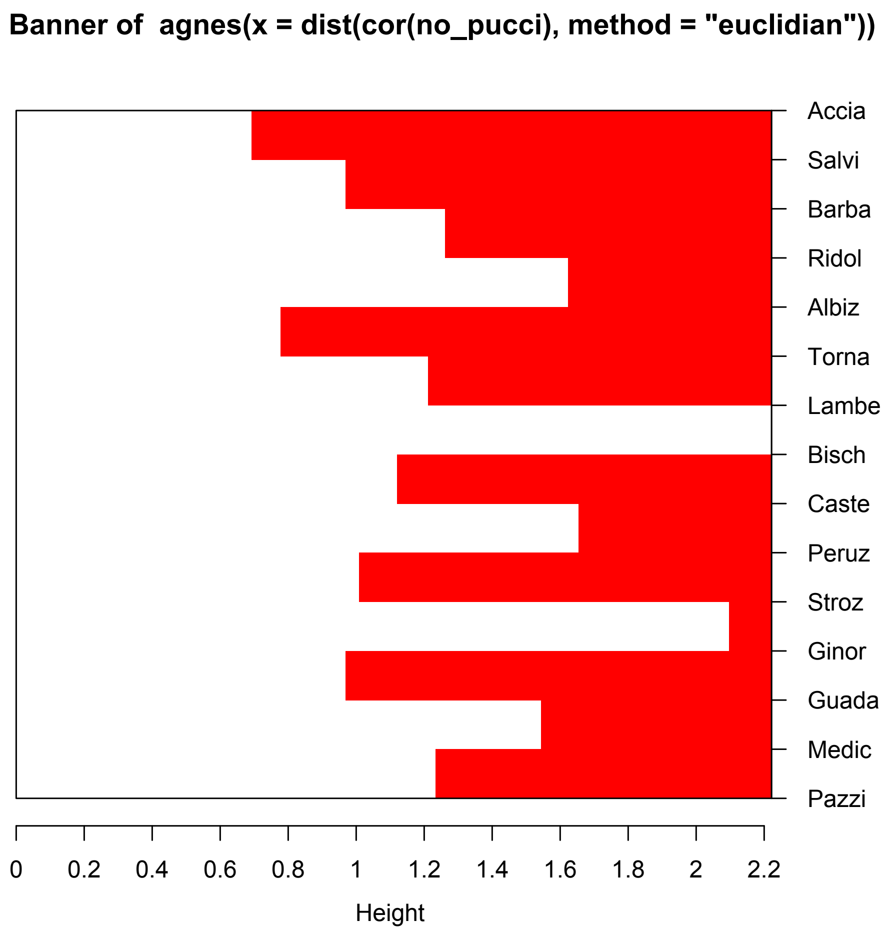




Take 2 minutes and discuss: How can you tell how many clusters there should be?

Hierarchical Clustering Can Be Based on Distance (left) or Correlation (right)



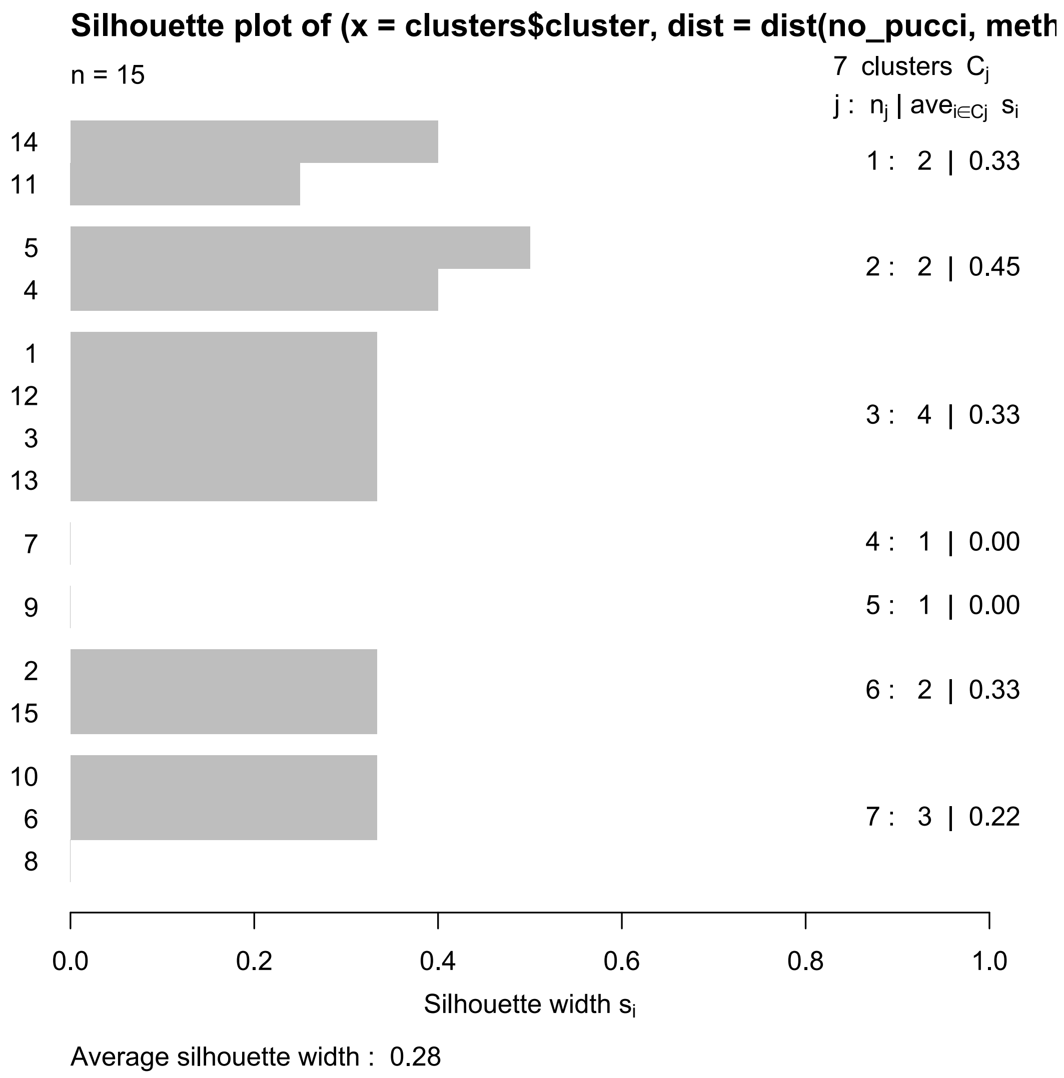


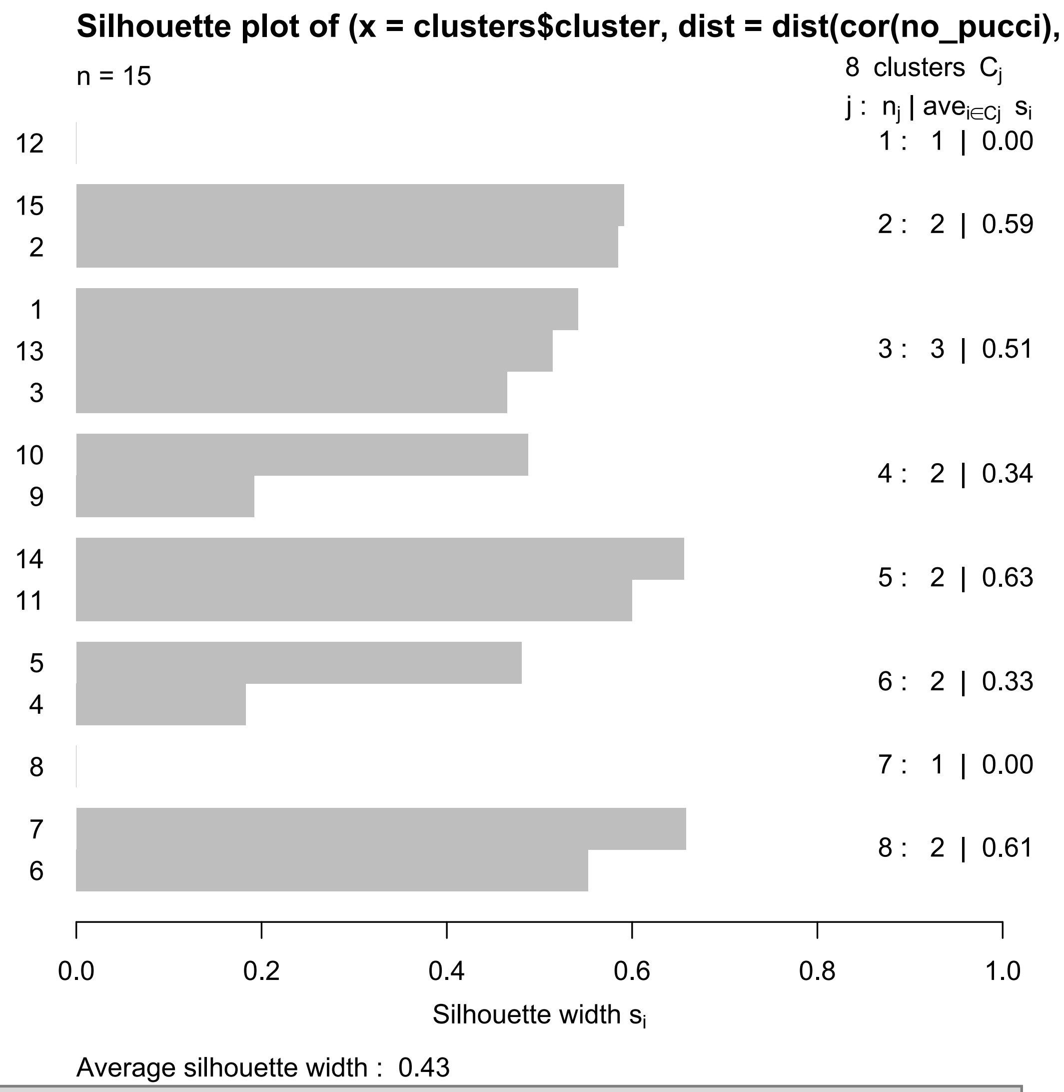
Take 2 minutes and discuss: How can you tell how many clusters there should be?
For banner plots, where the white lines stick into the red is where the clusters are made. Want clusters that are distinctive, so would like to make a cut that maximizes the information we get from the clustering.

k-*means* Clustering Clusters Iteratively around a Cluster Center

- First, choose how many clusters there should be (k)
- The algorithm will choose k observations to be the cluster's center
- The algorithm puts each observation into the cluster to whose center it is closest
- Then, recalculate what the center of each cluster should be and repeat the categorization process
- Iterate until no more observations switch clusters

k-*means* Clustering Clusters Iteratively around a Cluster Center (distance on left, correlation on right)



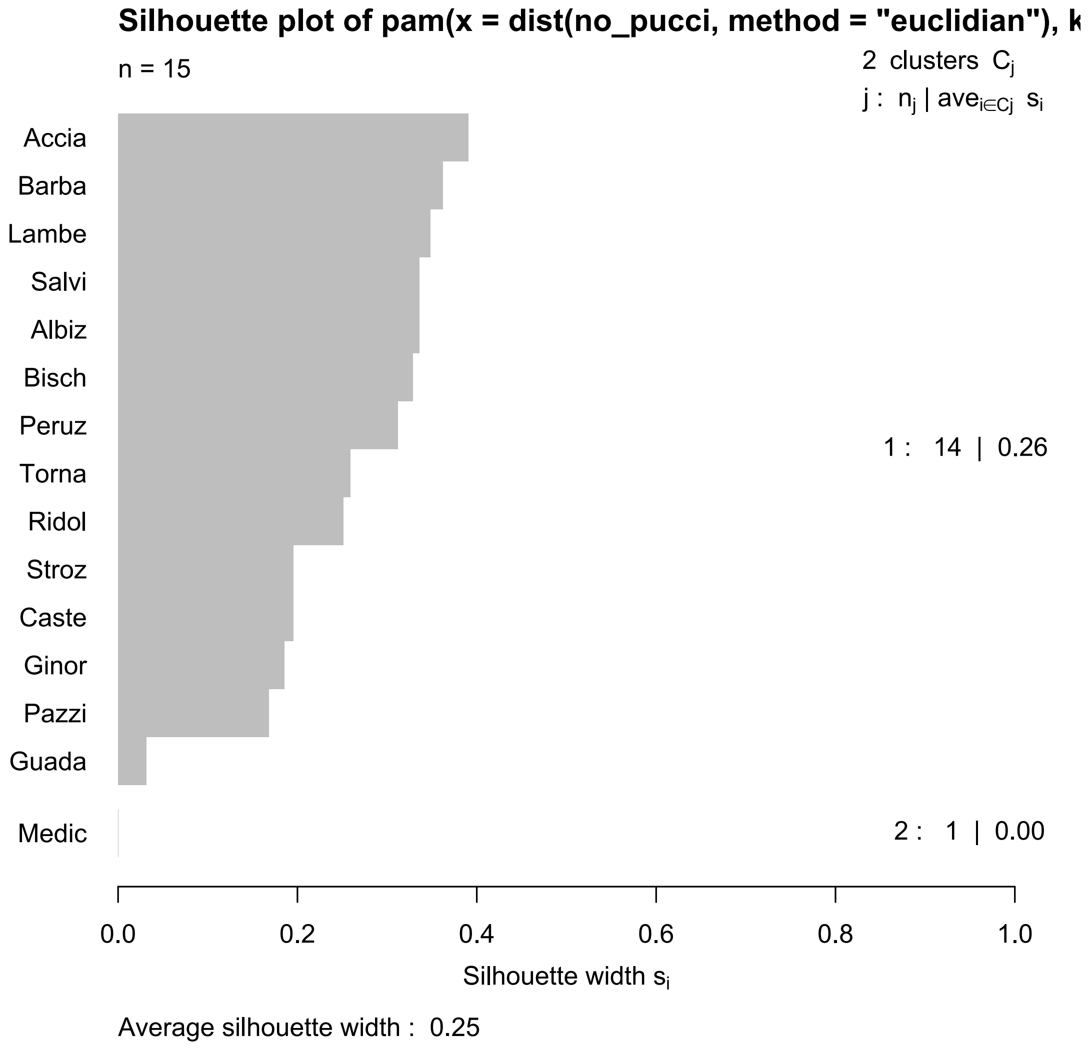


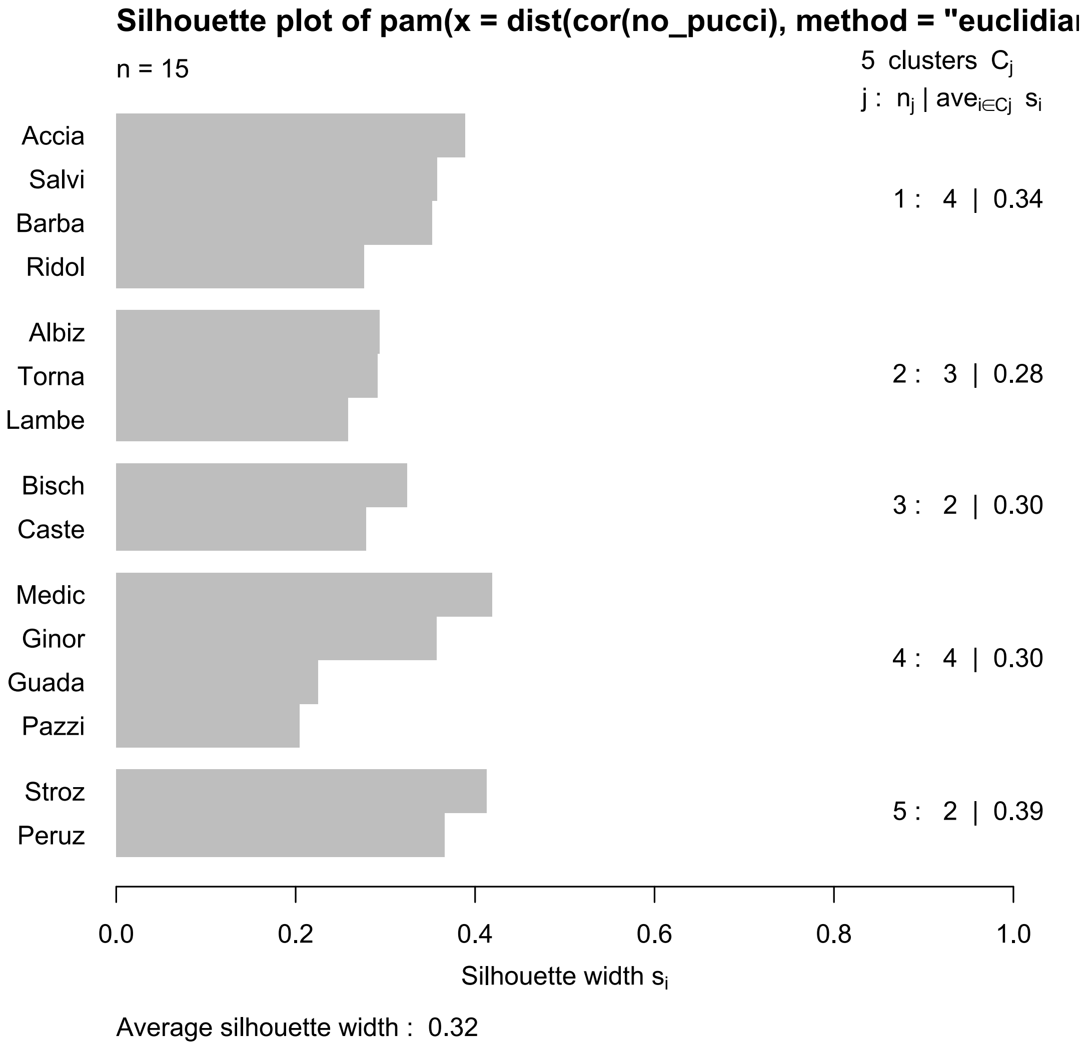
The average silhouette width of .43 is a little better than the distance clustering, but it isn't great.

Partitioning Around Medoids (PAM) takes a similar Approach as k-*means*

- Modernized version of k–*means* clustering
- Instead of recalculating cluster centers only after all observations have switched, perform these calculations constantly as each observation moves around
- The centers of the k -*means* clusters may not actually correspond to real observations
- PAM produces a medoid, or cluster center, that is useful representative of the objects in that cluster

Partitioning Around Medoids (PAM) takes a similar Approach as k-*means* (distance on left, correlation on right)

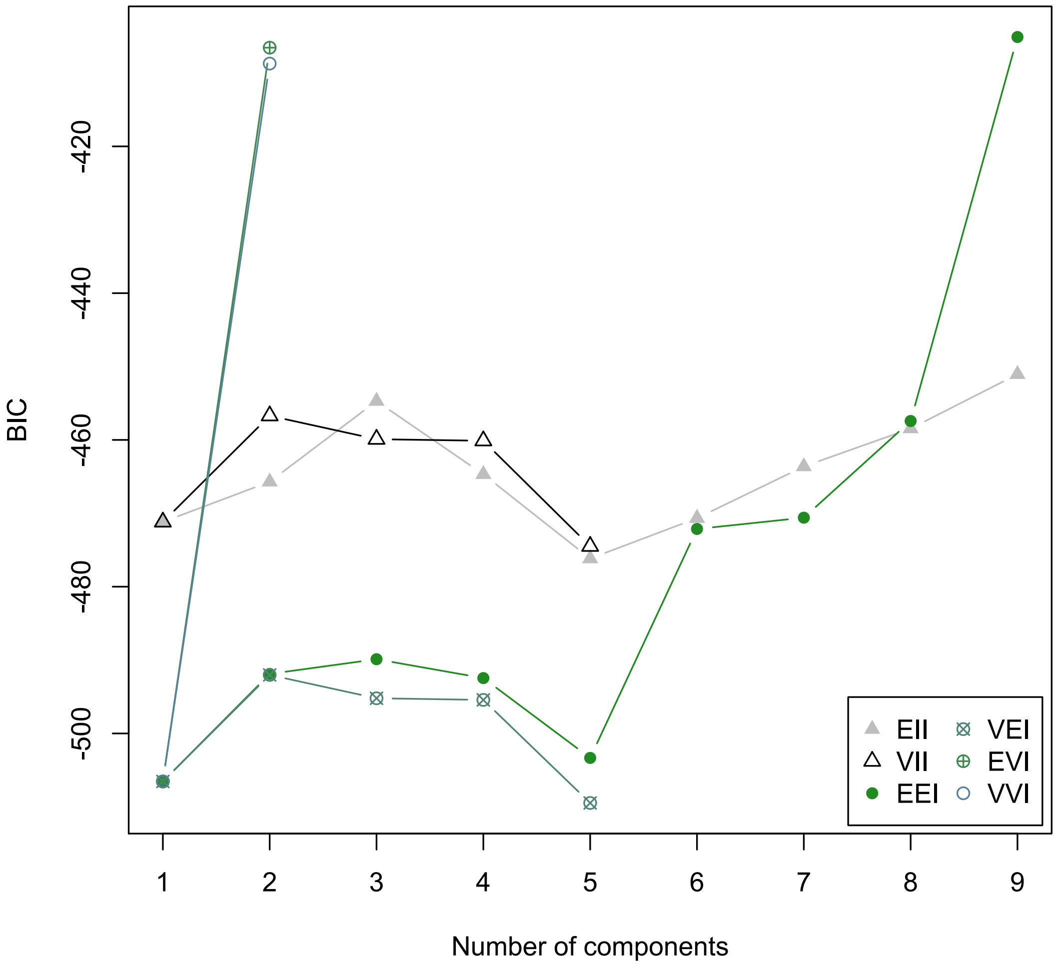


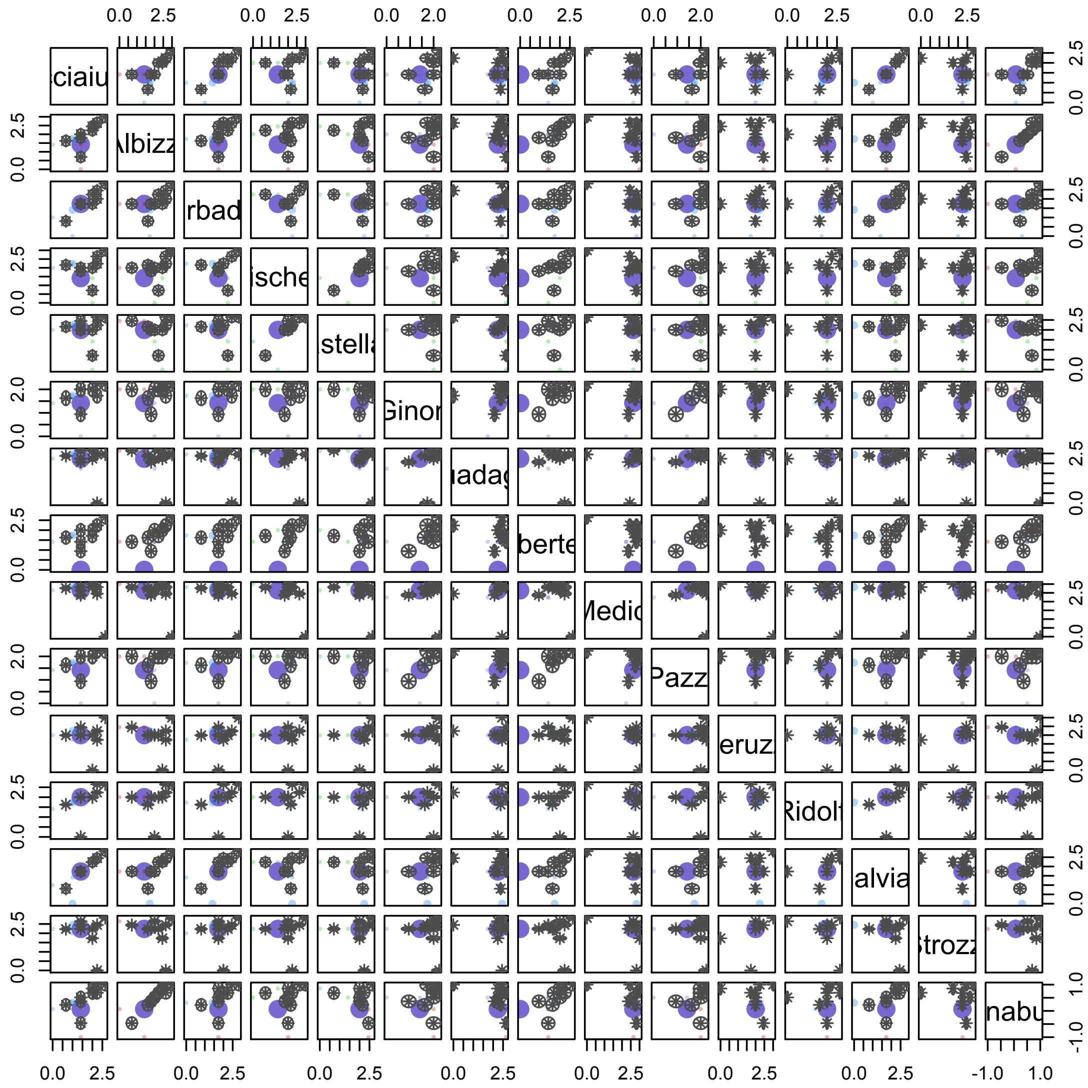


Model-based Clustering Takes a Maximum Likelihood Approach

- Instead of specifying clusters heuristically, allow clusters to arise probabilistically
- Usually assume that the clusters arise from a mixture of normal distributions
- Each cluster has a “weight” in the mixture that categorizes the data, obtained by an expectation-maximization algorithm
- This way, we can examine the relative fit of each clustering strategy to choose the most appropriate without having to make any decisions about the clusters beforehand
- Use information criterion that penalizes more clusters to gain the maximum distinction without over-fitting

Model-based Clustering Takes a Maximum Likelihood Approach



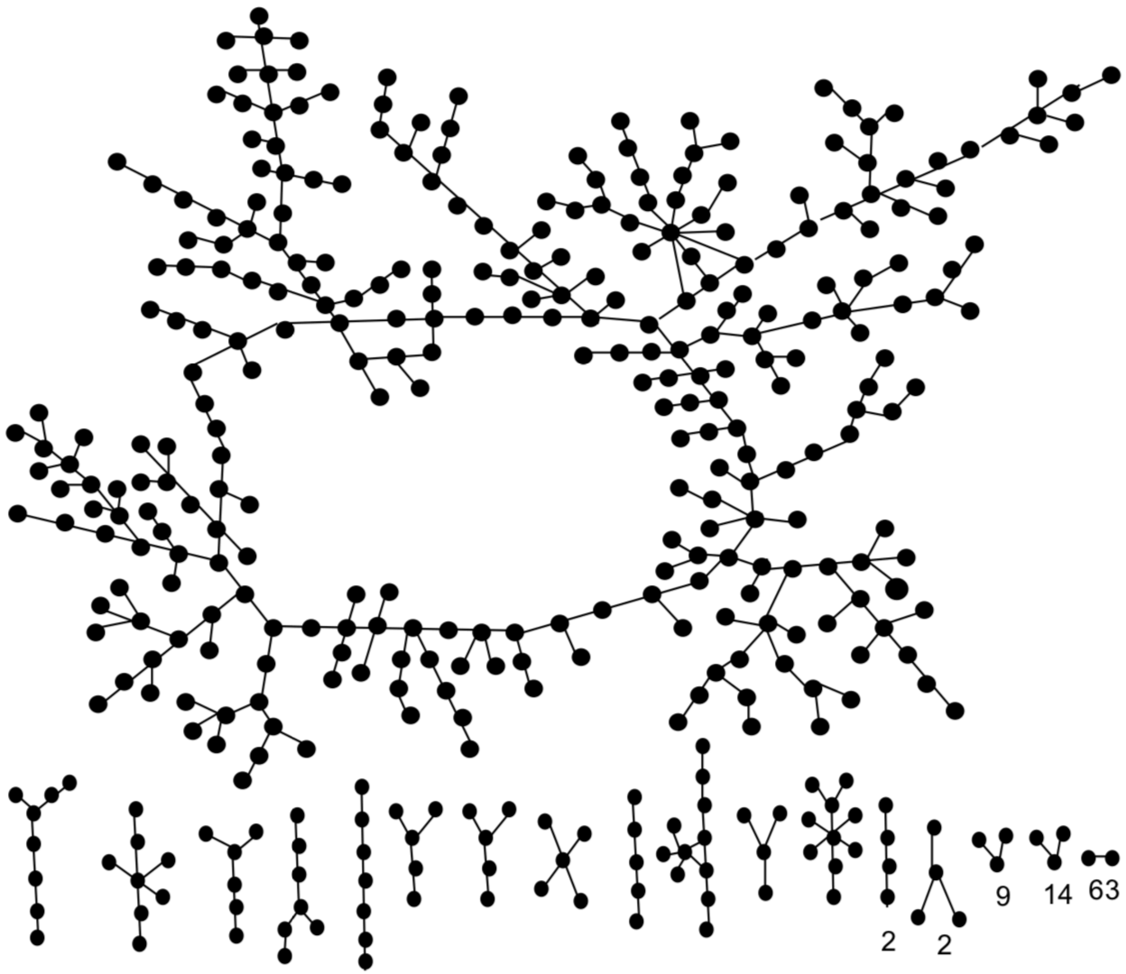


The BIC suggests that there should be 9 clusters of equal volume and shape. The model plot illustrates where the 9 clusters are fit in coordinate space for each run of the expectation maximization.

Clustering versus Core-periphery

Clustering Approach Tends to Think of Relatively Segmented Network Components

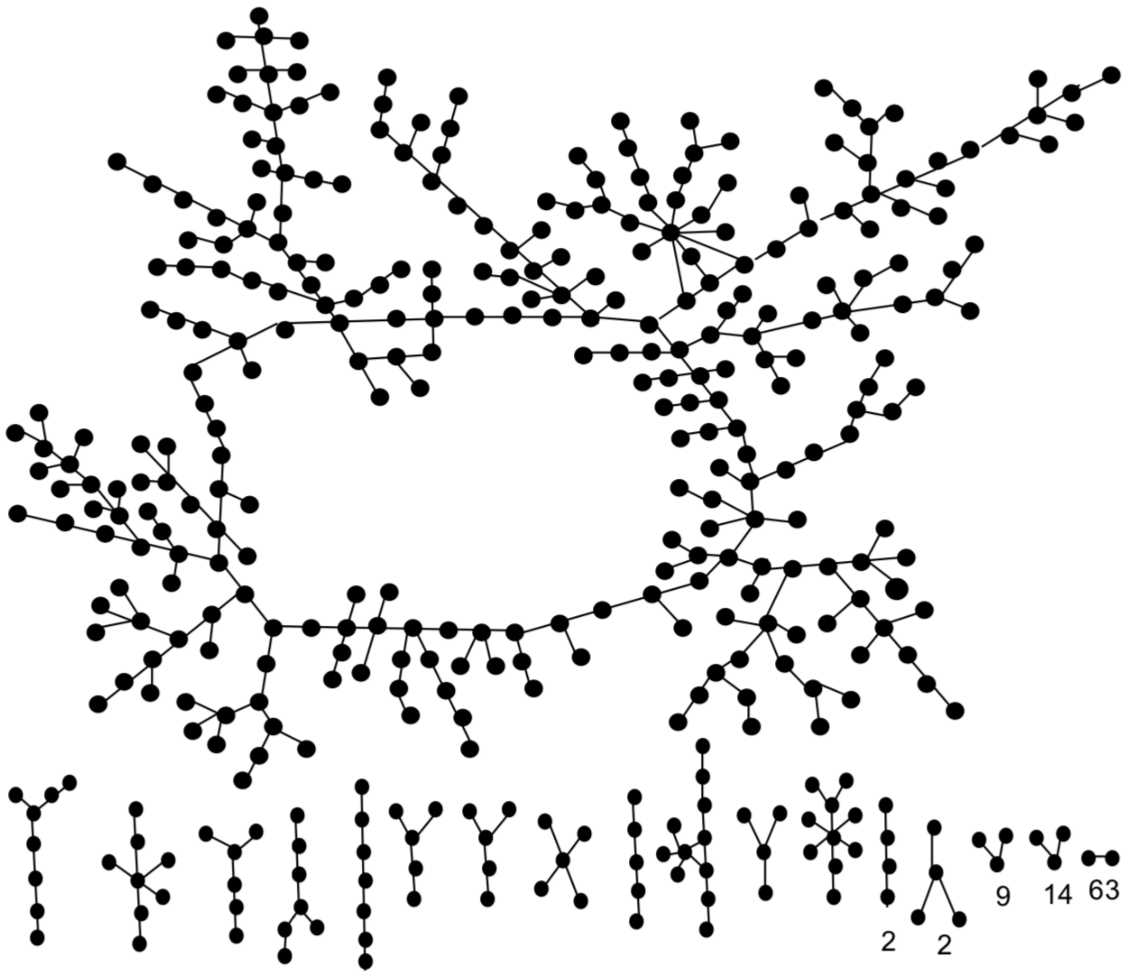
What if there was one large, “giant component” and a surrounding periphery?



This is a network of romantic relationships from a US high school. Take 2 minutes to discuss: why might there be a “giant component“ in this network?

Clustering Approach Tends to Think of Relatively Segmented Network Components

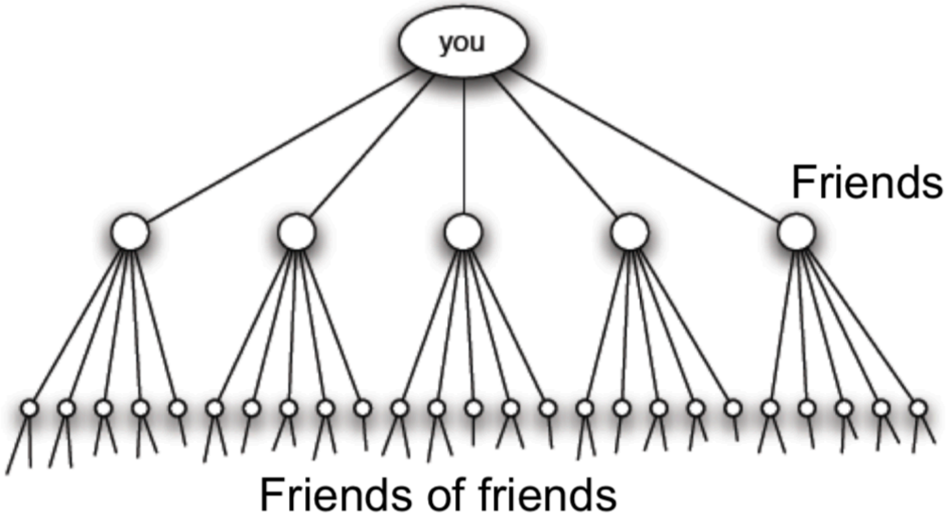
What if there was one large, “giant component” and a surrounding periphery?

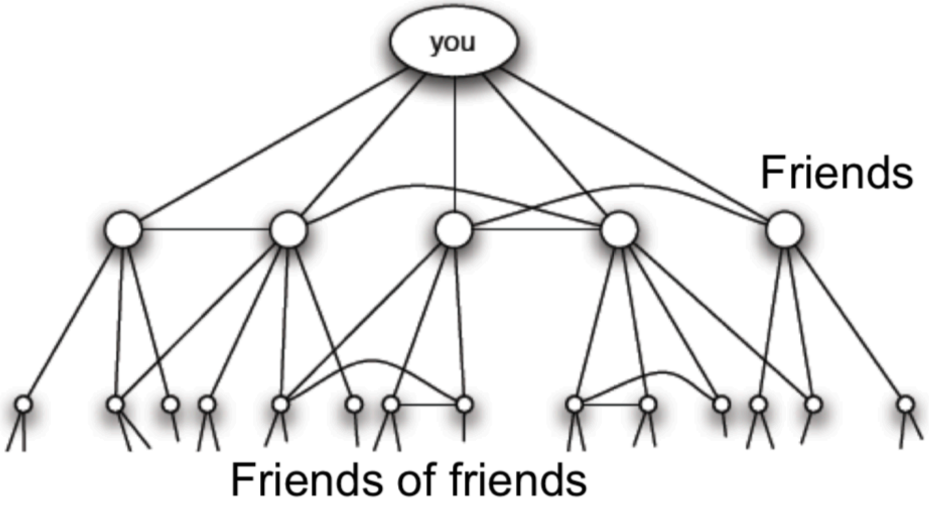


This is a network of romantic relationships from a US high school. Take 2 minutes to discuss: why might there be a “giant component“ in this network?

Only takes one relationship to connect a disconnected subgroup and make a giant component

Can think of This Connectivity as a Cascading Effect of Friends of Friends





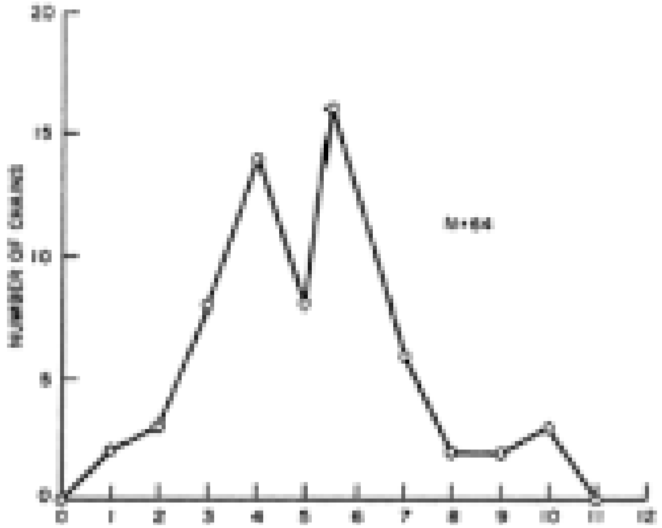
Even with interrelationships among friends, the total nodes connected becomes very large very fast.

The World Is Actually Quite Small

How many steps would it take for two people in the United States to get a message to each other?

Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

Length of chain, by number of steps



In 1967, study participants were able to relay a message from Nebraska to Boston via the US Postal Service in an average of about 5 steps.

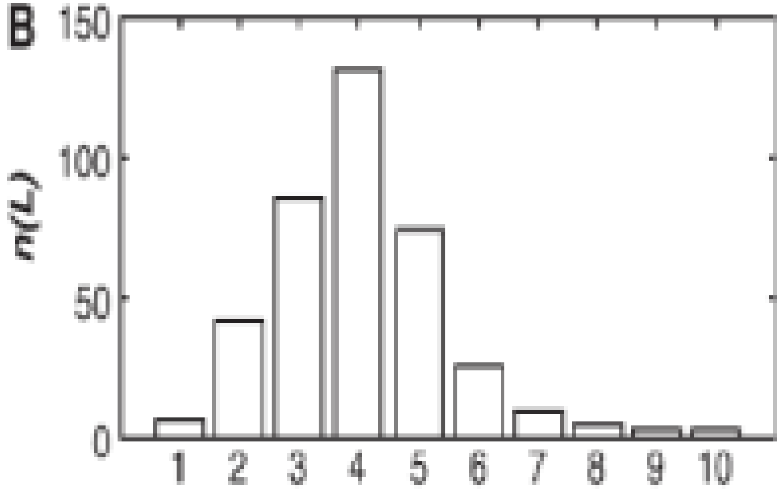
Source: Travers, J. & Milgram, S. "An Experimental Study of the Small World Problem." *Sociometry* 32, no. 4 (1969): 425-443.

What If We Conducted a Network Search Study Today?

How many steps would it take for two people in the United States to get a message to each other (if they had access to email)?

*We report on a global social-search experiment in which more than 60,000 e-mail users attempted to reach one of 18 target persons in 13 countries by forwarding messages to acquaintances. We find that **successful social search is conducted primarily through intermediate to weak strength ties**, does not require highly connected "hubs" to succeed, and, in contrast to unsuccessful social search, disproportionately relies on professional relationships. By accounting for the attrition of message chains, we estimate that social searches can reach their targets in a median of five to seven steps, depending on the separation of source and target, although small variations in chain lengths and participation rates generate large differences in target reachability. We conclude that although global social networks are, in principle, searchable, actual success depends sensitively on individual incentives.*

Length of chain, by number of steps

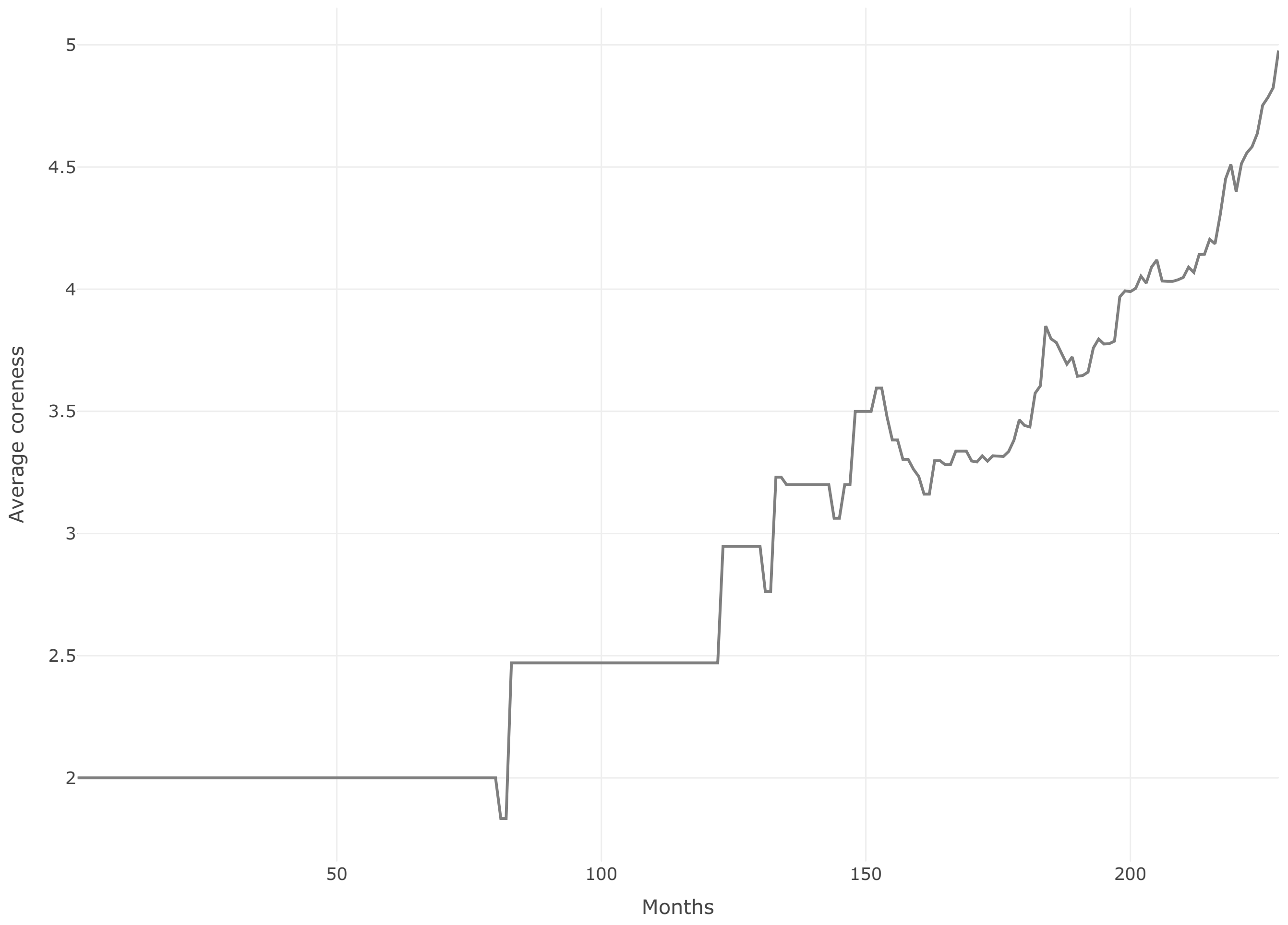


In 2003, study participants were able to relay a message to over a dozen countries via email in an average of about 5 steps. In both studies, successful search was routed through weak ties.

Source: Dodds, P.S., Muhamed, R., & Watts, D.J. "An experimental study of search in global social networks." *Science* 301, no. 8 (2003): 827-829.

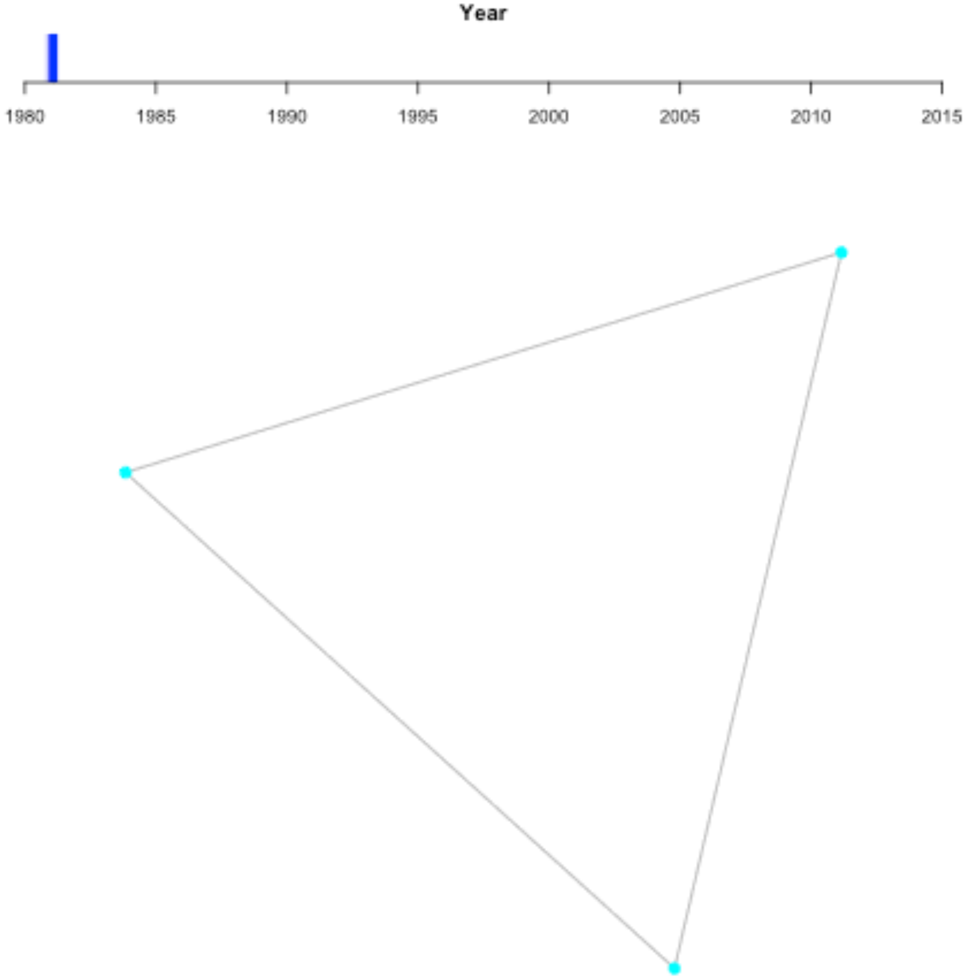
Core Periphery Structure in Venture Capital

Does "Core-ness" of the Network Increase Over Time?



*The plot charts the average coreness, in terms of the highest k-core they are a part of, of **every** member of the network (not just the core itself). The data start around 1980, so coreness really starts to pick up around the late 80s and early 90s.*

Can Illustrate this Structure in the Changing Shape of the Network Over Time



The network develops a densely connected center and a surrounding periphery towards the late 1980s and onward into the 1990s.

Small Worlds in the Actor Network Using the Oracle of Bacon and Neo4j

The Oracle of Bacon at Virginia

Mapping the Hollywood actor network at http://oracleofbacon.org/

Out of 2,331,812 linkable actors, the average number of steps to create a chain to Kevin Bacon is 3.029. Almost 3500 actors can be linked in one step, and over 400,000 in two steps.

Source: http://oracleofbacon.org/onecenter.php?who=Kevin+Bacon

The Oracle of Bacon at Virginia

Mapping the Hollywood actor network at http://oracleofbacon.org/

01 — What hypotheses guided your attempts to find high Bacon numbers?

02 — What was the highest Bacon number you found? What are the links in the chain?

03 — What makes the world so small? Are there links appear in your chains very frequently?

Out of 2,331,812 linkable actors, the average number of steps to create a chain to Kevin Bacon is 3.029. Almost 3500 actors can be linked in one step, and over 400,000 in two steps.

Source: http://oracleofbacon.org/onecenter.php?who=Kevin+Bacon

Who Might Be More Central than Kevin Bacon?

In fact, Kevin Bacon is only the 455th most central actor in the Hollywood network

Source: http://oracleofbacon.org/center_list.php

Who Might Be More Central than Kevin Bacon?

In fact, Kevin Bacon is only the 455th most central actor in the Hollywood network

1
Eric Roberts
2.83

2
Michael Madsen
2.86

3
Samuel L. Jackson
2.87

4
Danny Trejo
2.87

5
Willem Dafoe
2.88

6
Harvey Keitel
2.88

Source: http://oracleofbacon.org/center_list.php

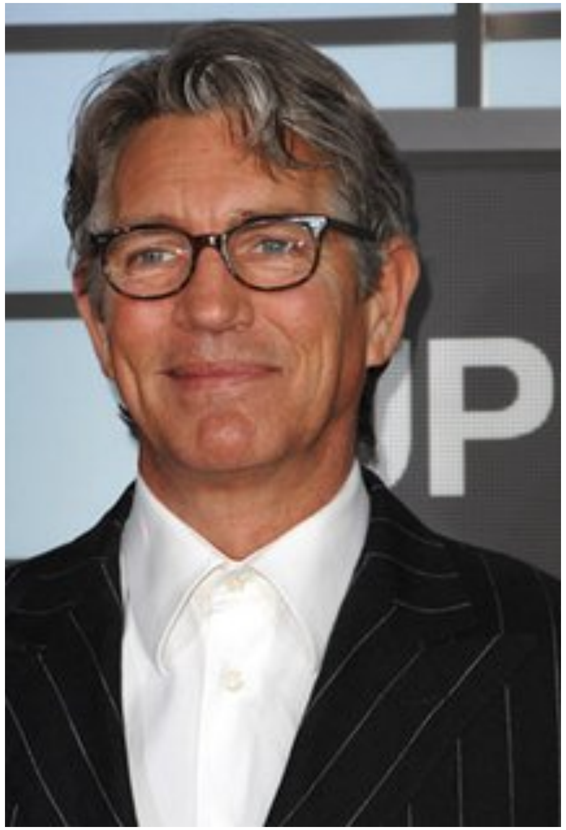
Why Is Eric Roberts So Central?
Tracing the origin of his network

"After appearing in such daytime soaps as Another World (1964) and How to Survive a Marriage (1974), his career began to shift fast forward when he copped a leading role in a major film"

"A wide range of fascinating, whacked-out roles were immediately offered to him on a silver plate"

"Soon began appearing as flashy secondary villains and creepies that showcased other stars instead"

"His film career began to slide in the late 1980s, appearing in more quantity than quality pictures"



Eric Roberts has 472 (!) acting credits as of 2017. He has starred in soap operas, blockbuster hits, critically-acclaimed films, and universally-panned busts. This wide variety of roles results in an incredibly diversified network of connections to other actors.

Source: http://www.imdb.com/name/nm0000616/bio?ref_=nm_ov_bio_sm

Mapping the Actor Network in Neo4j

- Quick intro to working with and visualizing network data in Neo4j
- Navigate to https://neo4j.com/download/other-releases/ and download and install the community edition (can set up desktop later)
- Download the file and follow the installation setup

Once you are in the browser, open up and take a look at the code "neo4j movie script.txt" on Canvas. It will let you explore a subset of the movie data as a graph object in Neo4j.

Questions?

Next class, Tuesday October 31

Topic

Representing relationships between multiple networks

Details

In this session, we will draw relationships between completely separate networks. In this way, it is possible to build affiliation, or bipartite, networks to relate people to products, groups, events, and so on. We will extend the individual, group, and network level measures to these multiple networks to understand how their properties can govern multiple relationships at once.

Reading

Easley & Kleinberg, Ch. 4