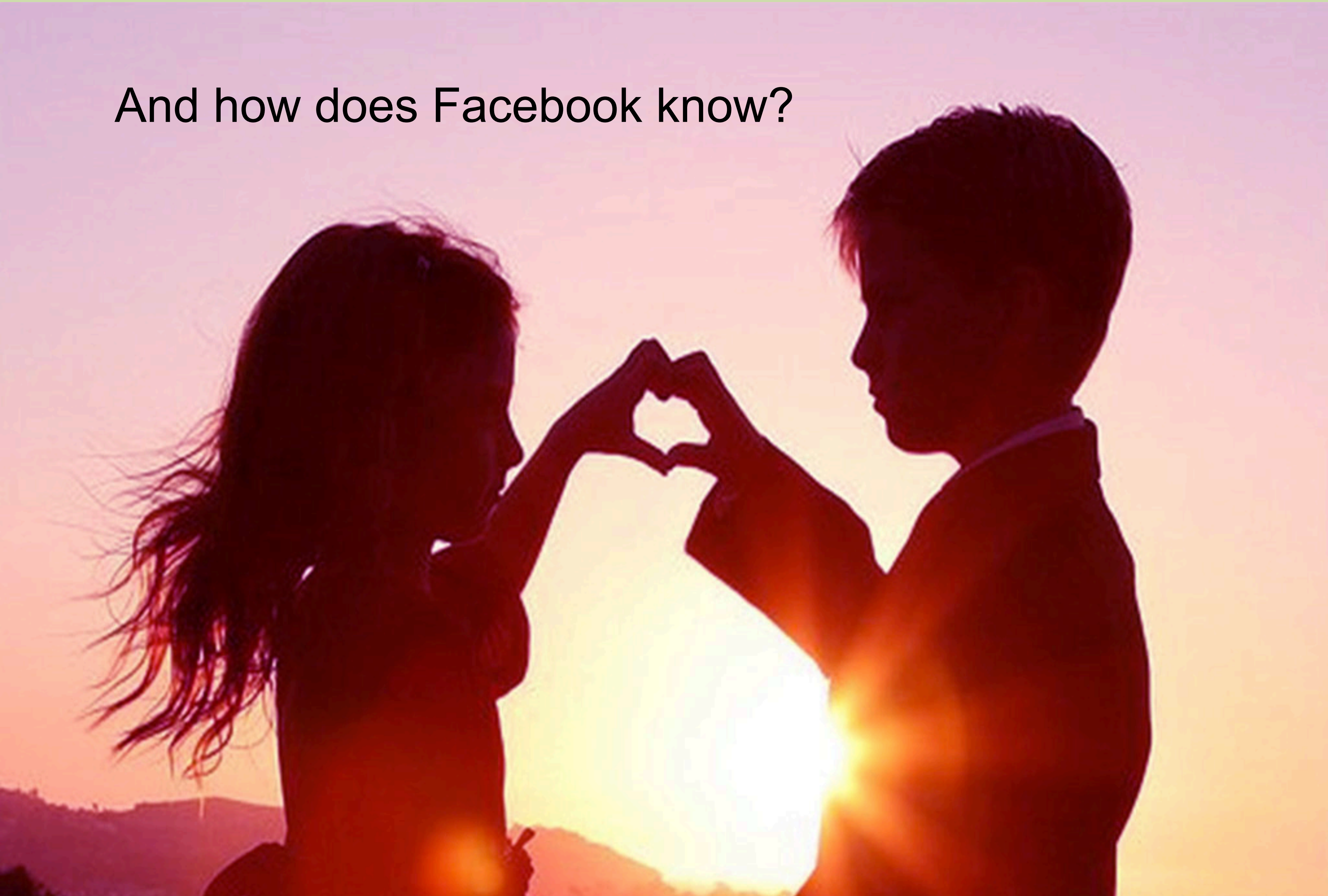
Graphs

Chris Piech

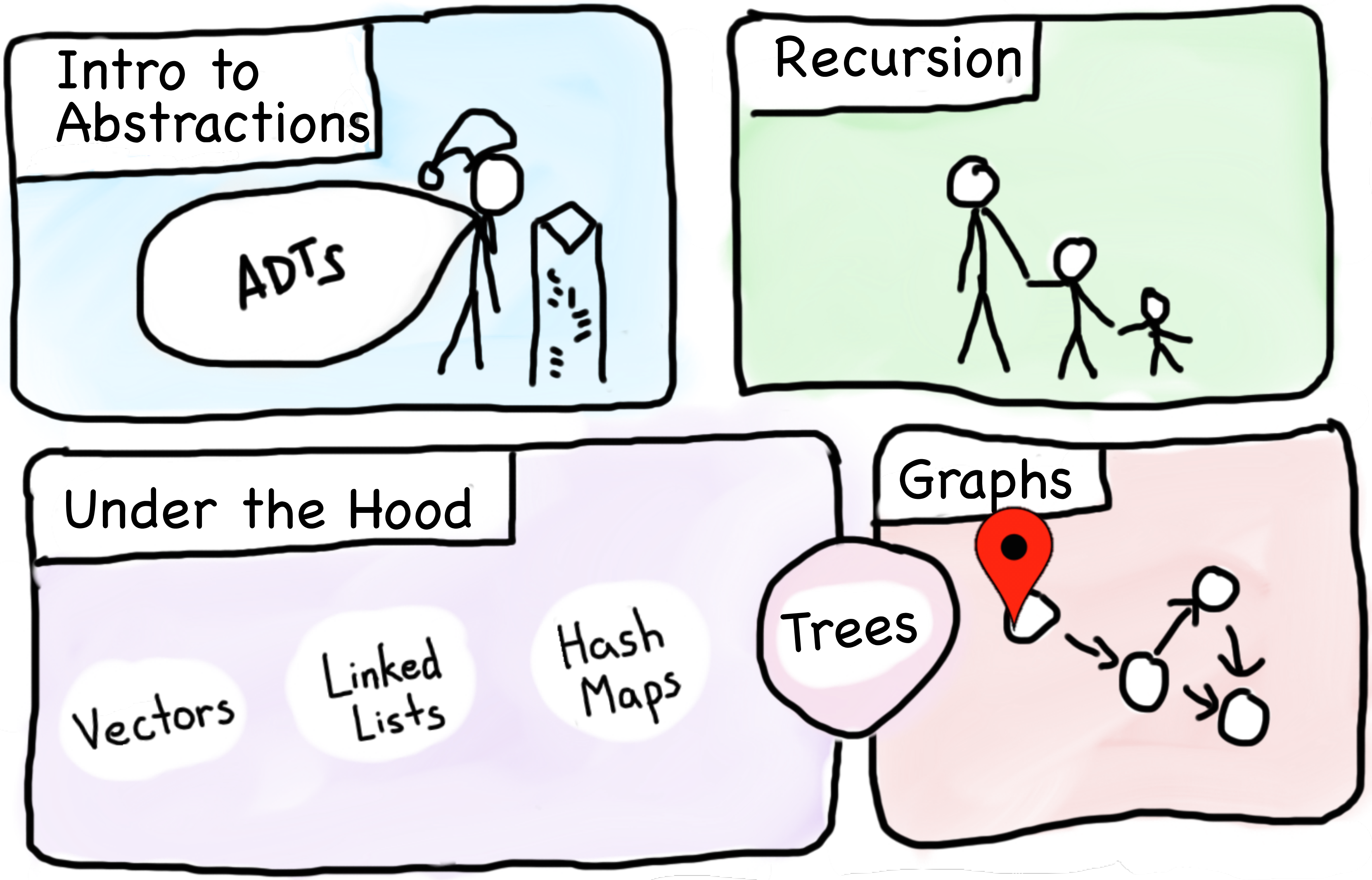
CS 106B
Lecture 20
Feb 24, 2016

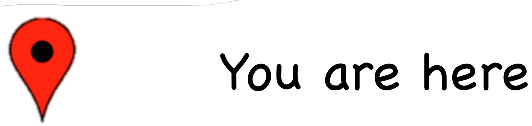
Who Do You Love

And how does Facebook know?



Course Syllabus

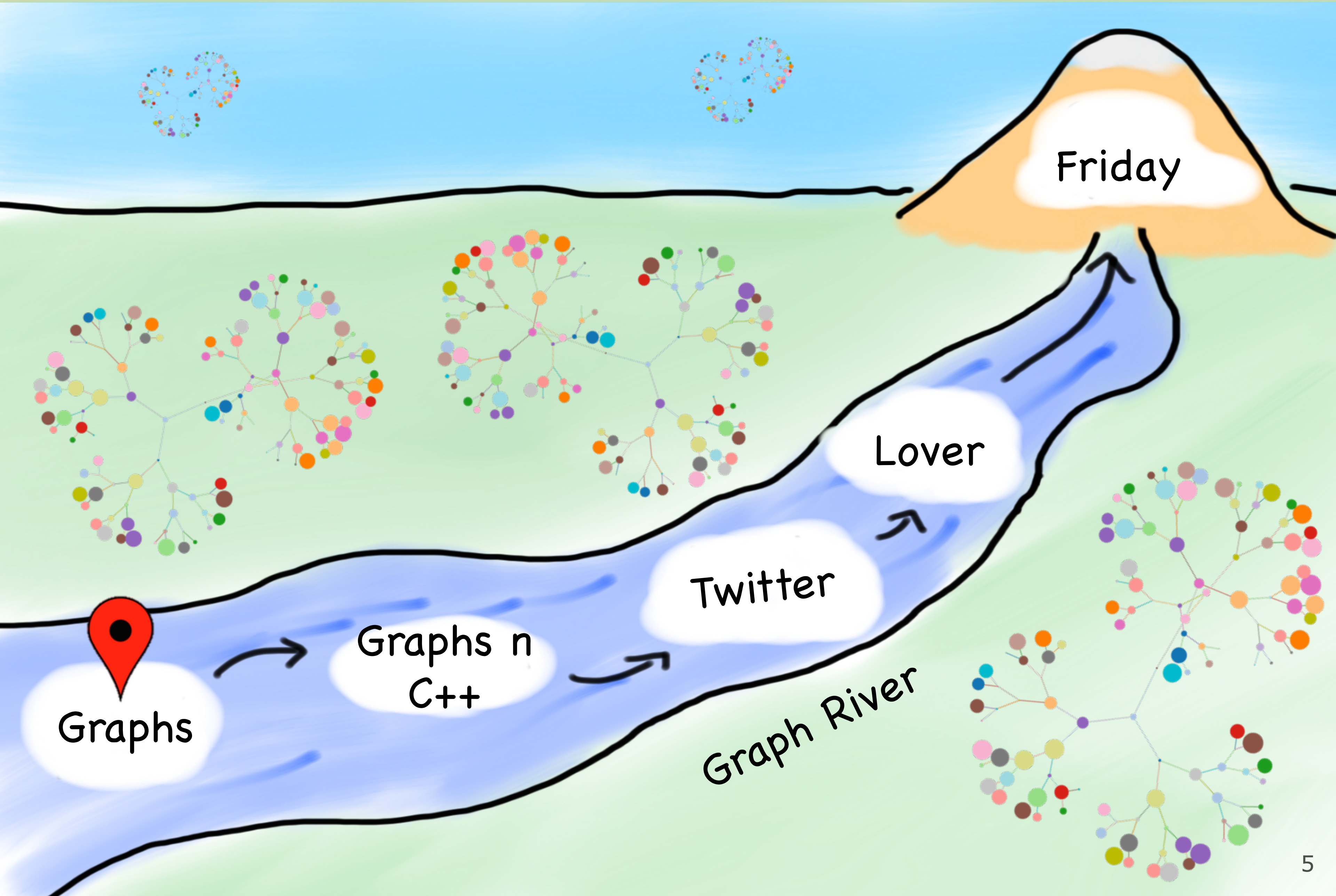




Today's Goal

1. Introduction to Graphs
2. Graphs in C++

Today's Route



Tree Definition

Only One Parent

No Cycles


WHAT IF I TOLD YOU
THERE ARE NO "RULES"

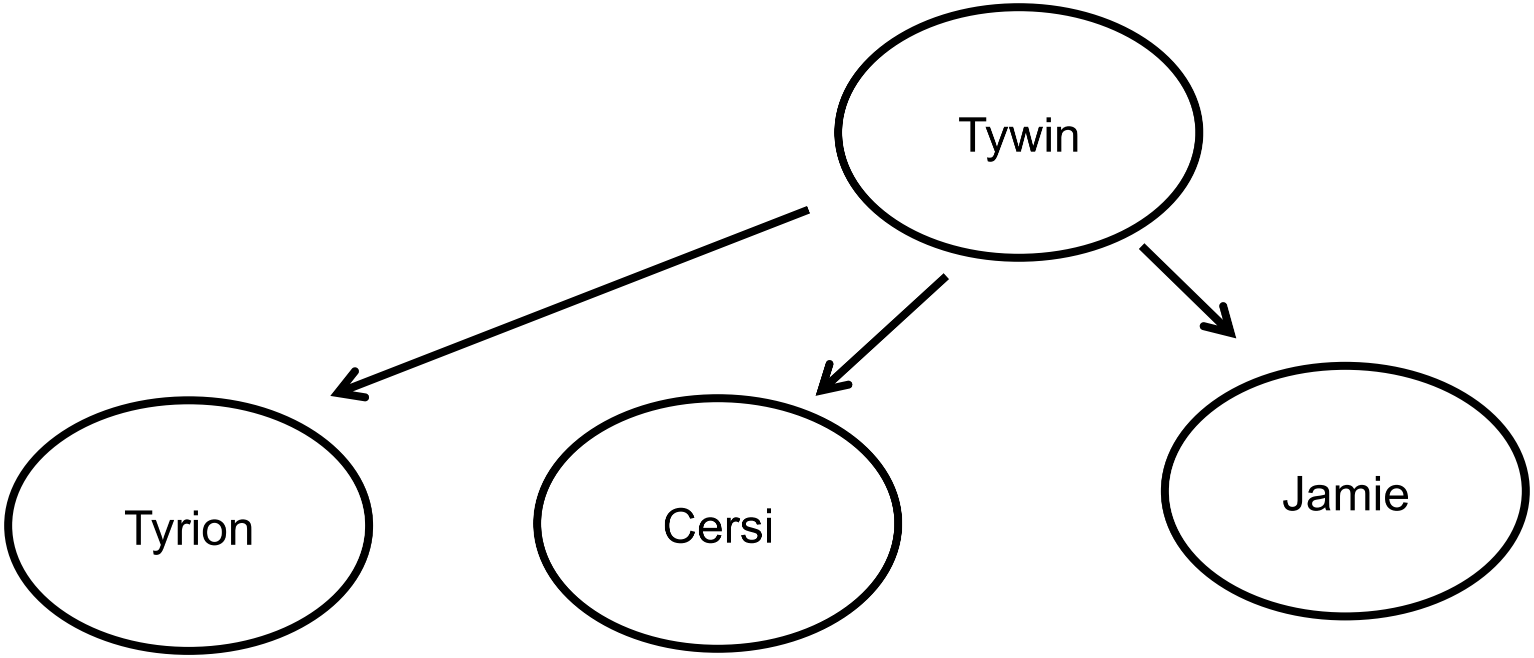
Graph Definition

A **graph** is a mathematical structure for representing relationships using nodes and edges.

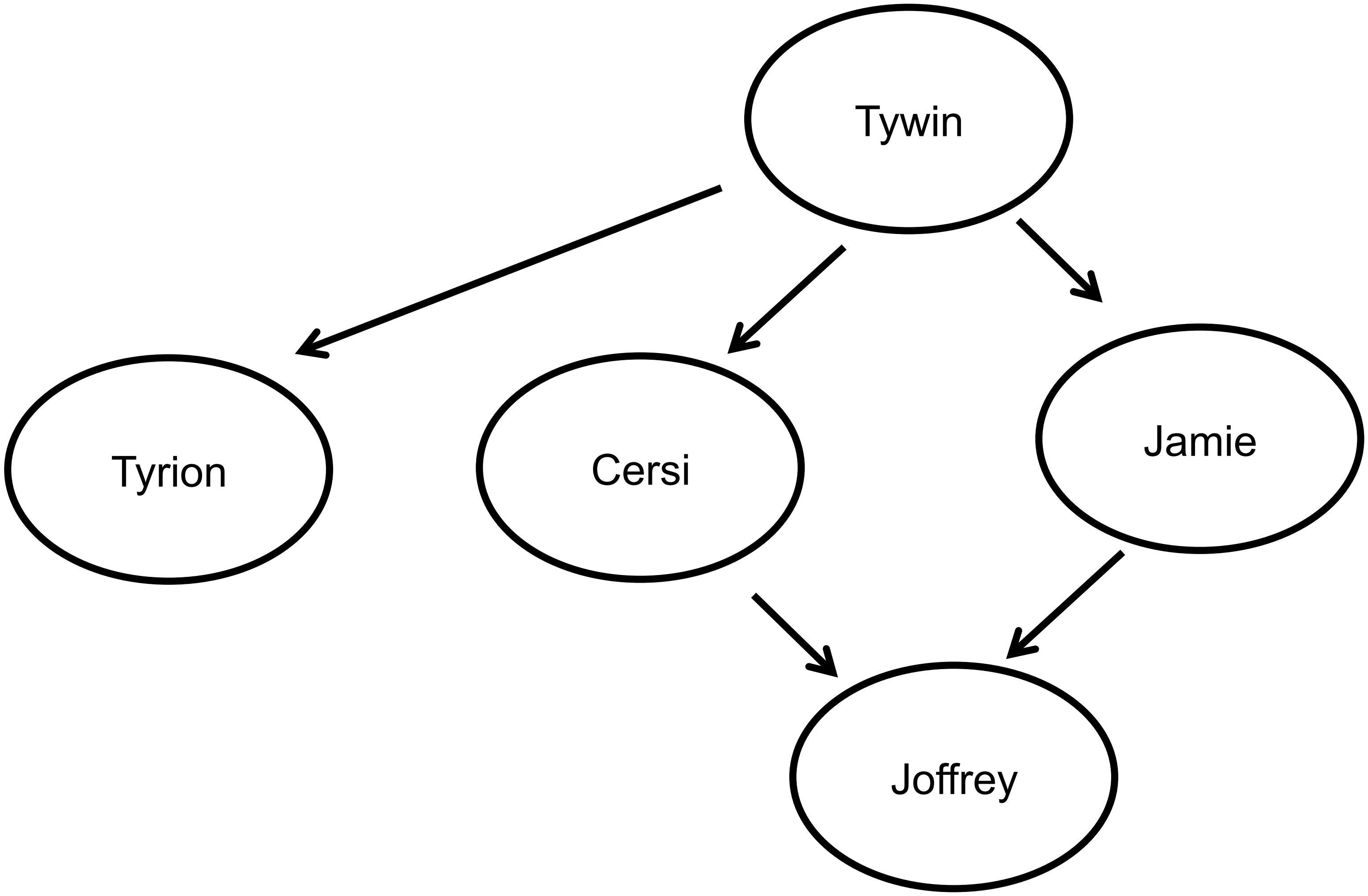
*Just like a tree without the rules


We can have a family tree?

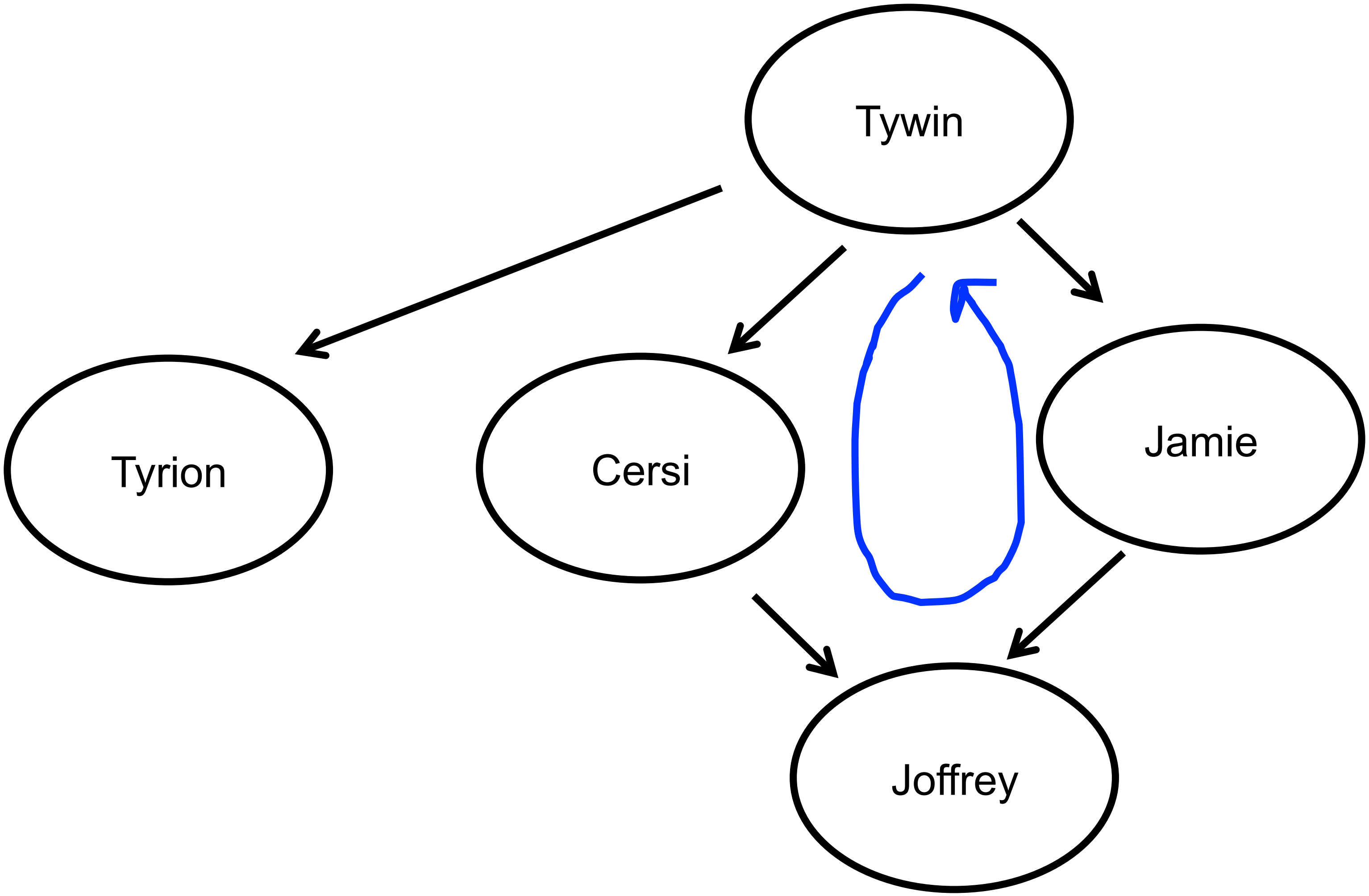
Not a Tree



Not a Tree



Not a Tree




graph
We can have a family ~~tree~~?

Simple Graph

```
struct Node{
   string value;
   Vector<Edge *> edges;
};

struct Edge{
   Node * start;
   Node * end;
};

struct Graph{
   Set<Node *> nodes;
   Set<Edge*> edges;
};
```

Simple Graph

```
struct Node{
  string value;
  Vector<Edge *> edges;
};
struct Edge{
  Node * start;
  Node * end;
};
struct Graph{
  Set<Node *> nodes;
  Set<Edge*> edges;
};
```

We allow for more interesting edges

Simple Graph

```
struct Node{
   string value;
   Vector<Edge *> edges;
};

struct Edge{
   Node * start;
   Node * end;
};

struct Graph{
   Set<Node *> nodes;
   Set<Edge*> edges;
};
```

We allow for more interesting edges

Simple Graph

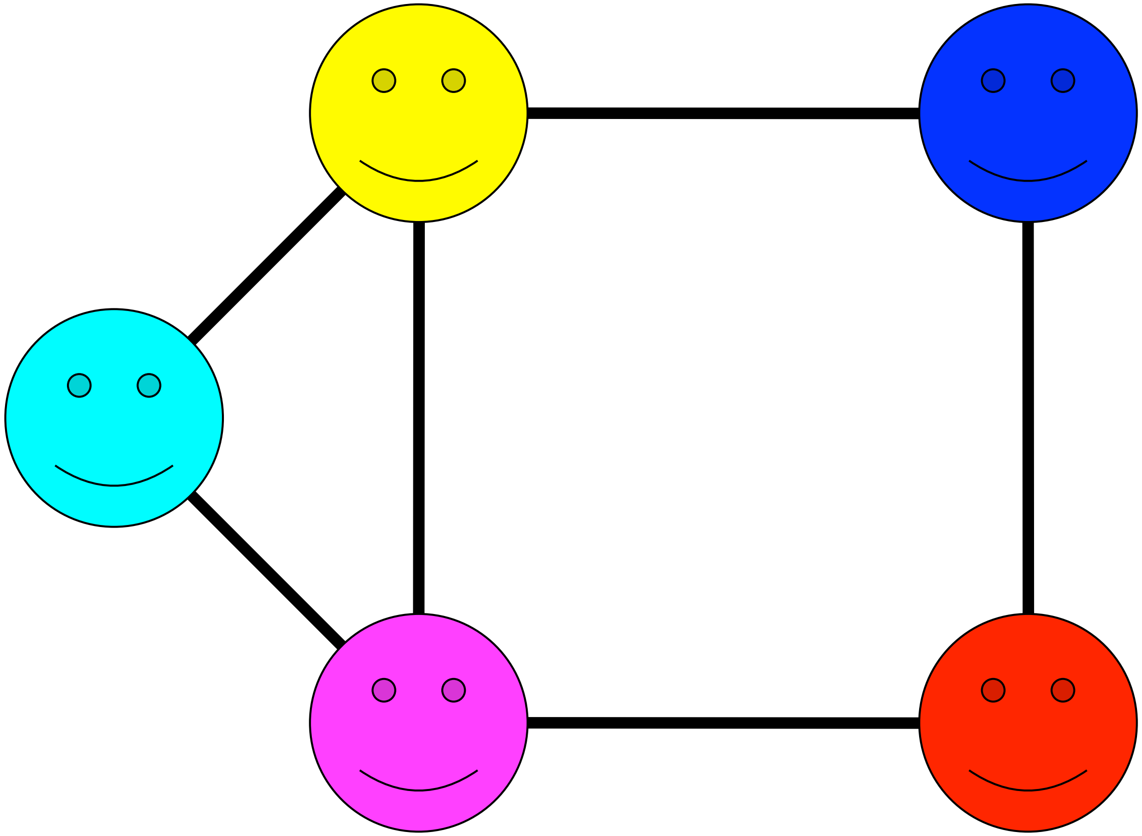
```
struct Node{
   string value;
   Vector<Edge *> edges;
};
```

```
struct Edge{
   Node * start;
   Node * end;
   double weight;
};
```

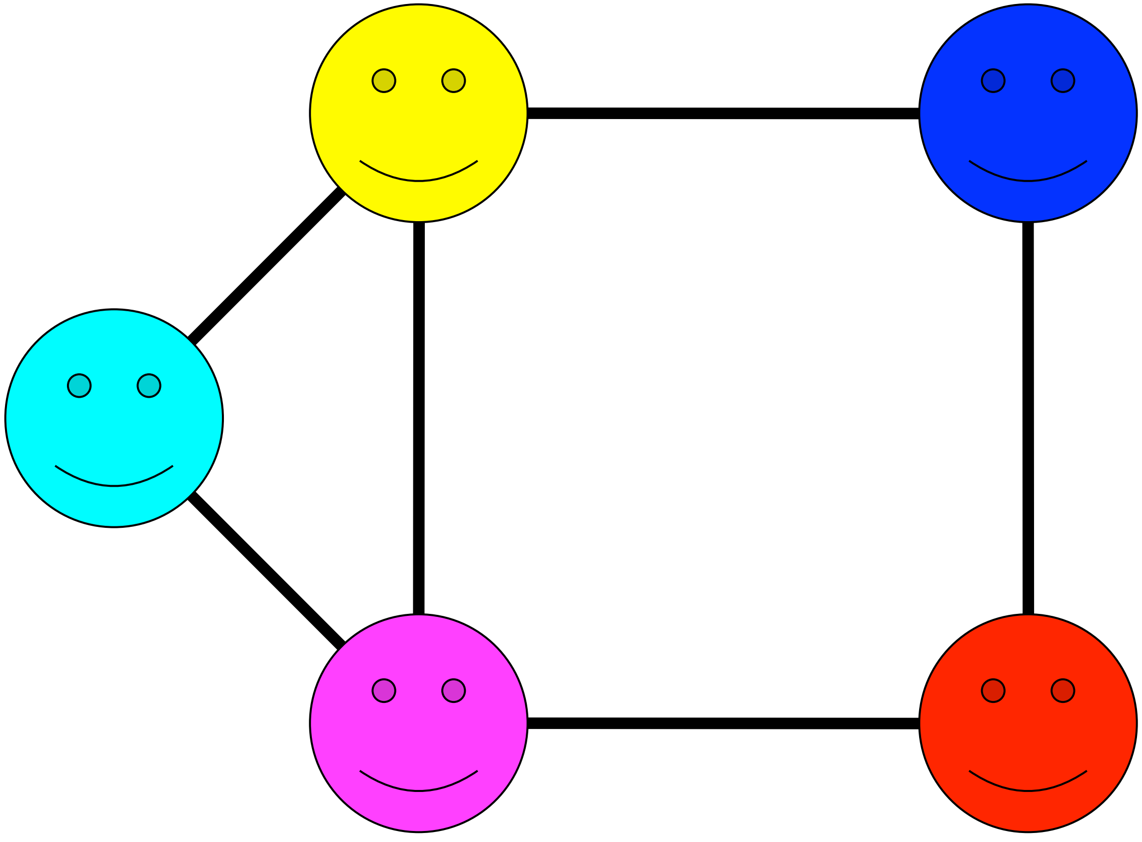
We allow for more interesting edges

```
struct Graph{
   Set<Node *> nodes;
```

Simple Graph

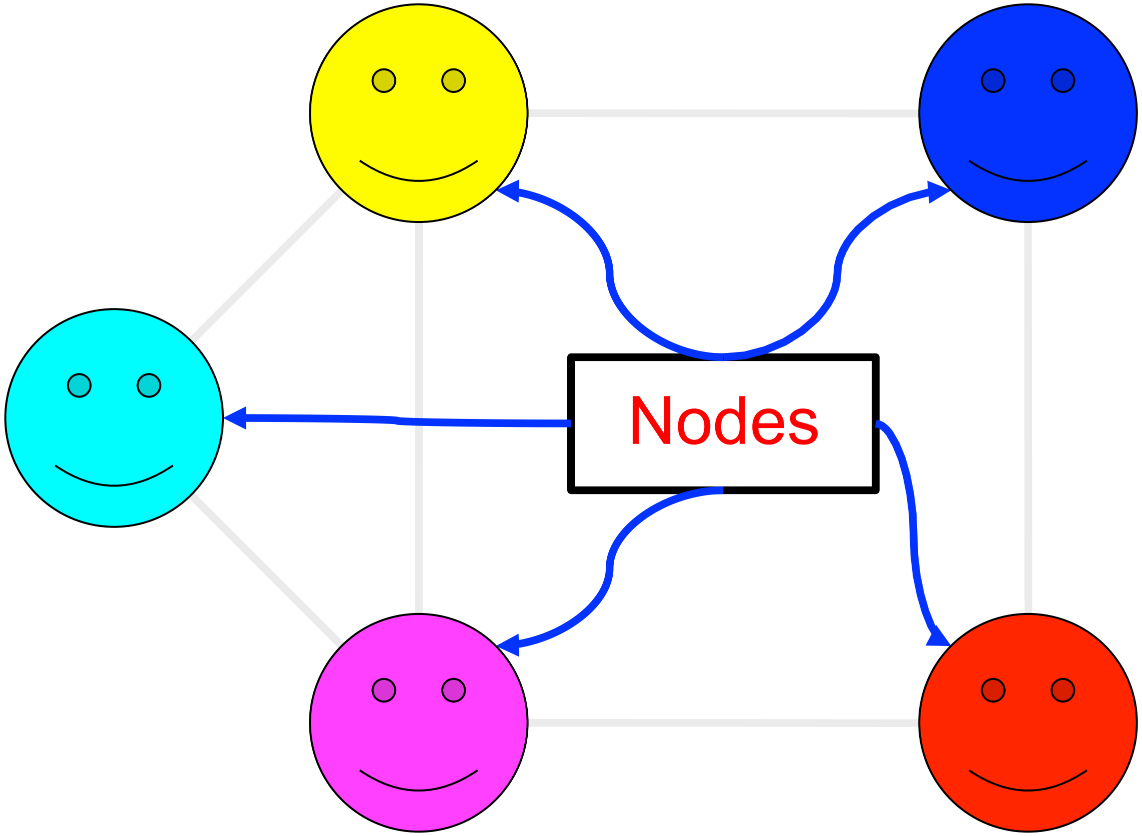


Simple Graph



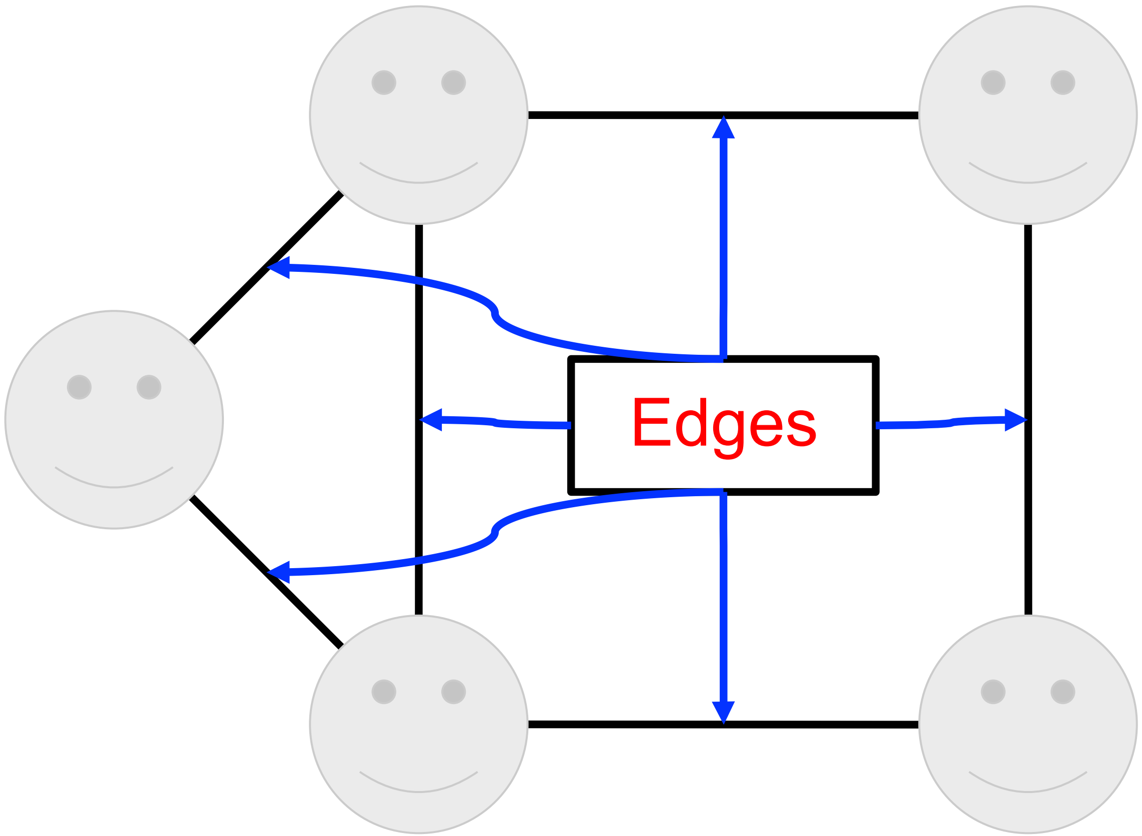
A graph consists of a set of **nodes** connected by **edges**.

Graph Nodes



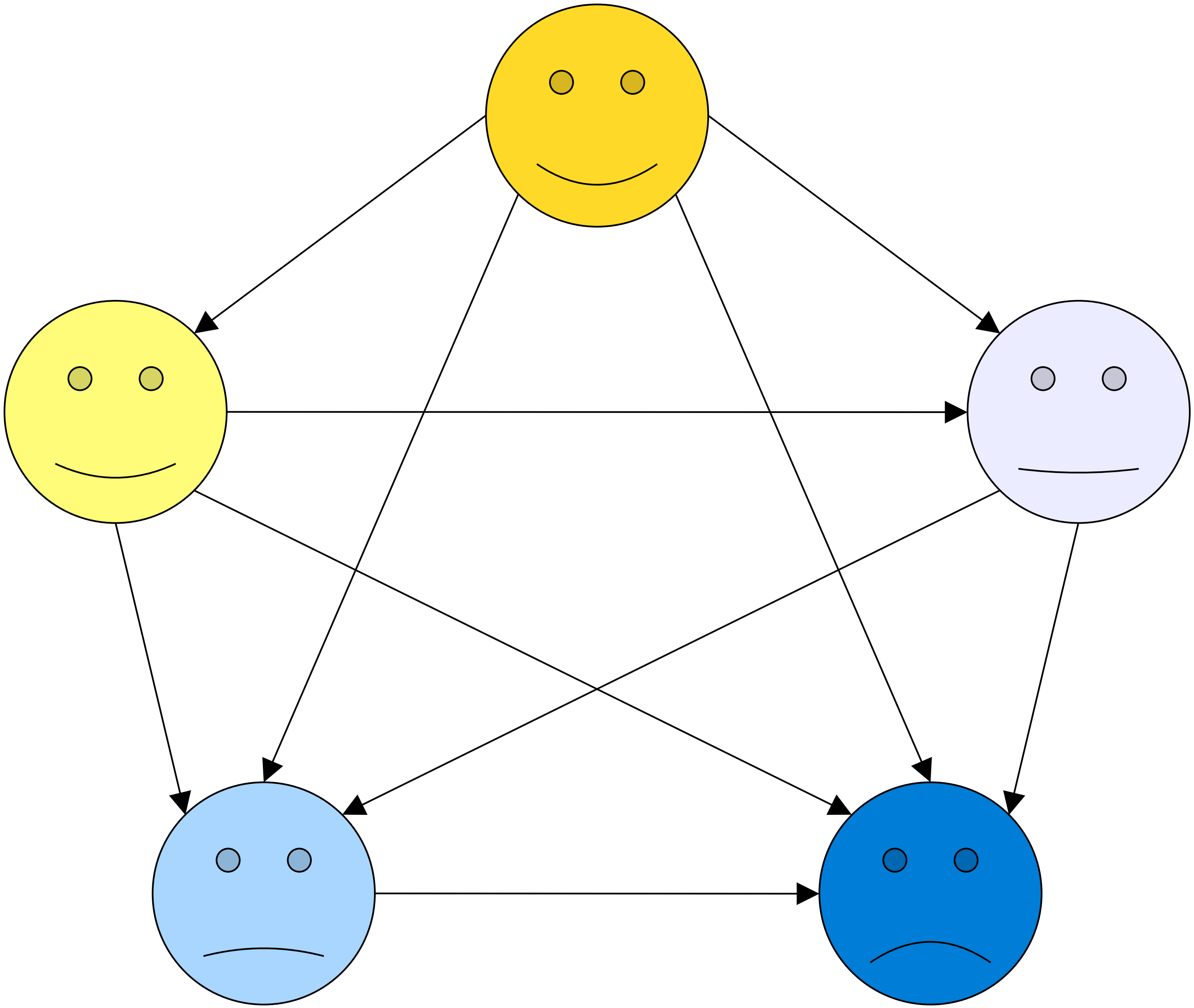
A graph consists of a set of **nodes** connected by **edges**.

Graph Edges

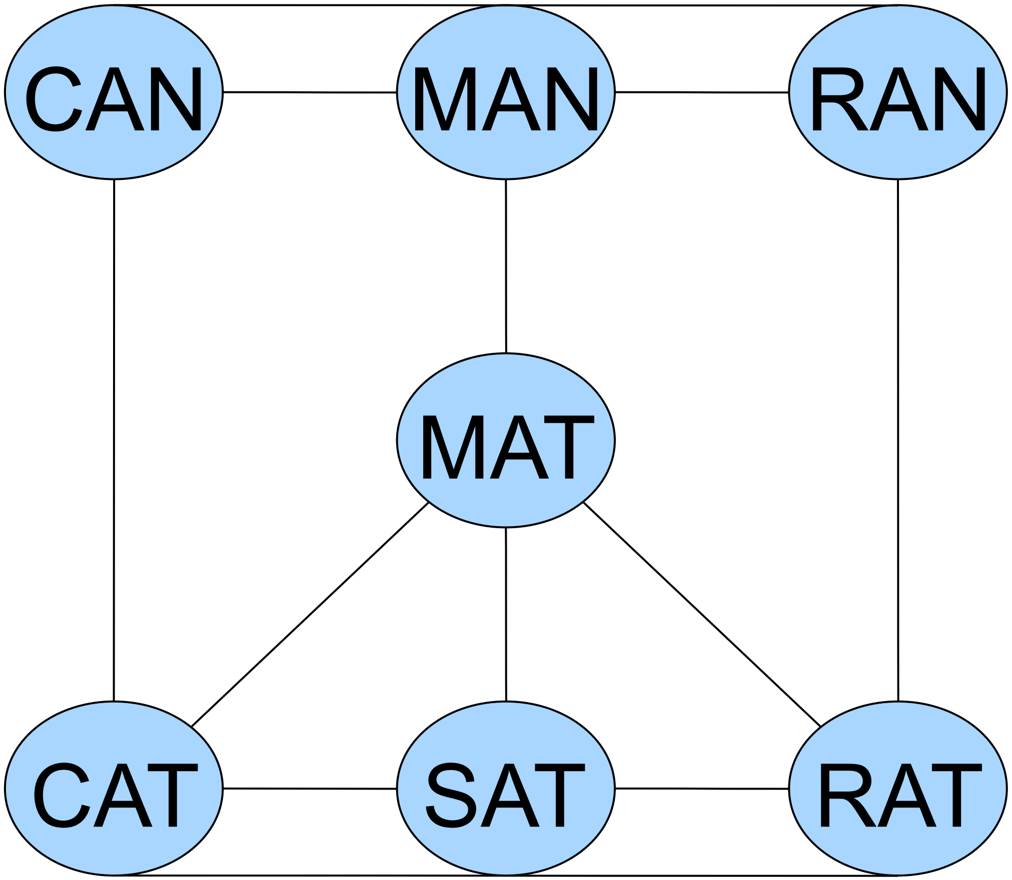


A graph consists of a set of **nodes** connected by **edges**.

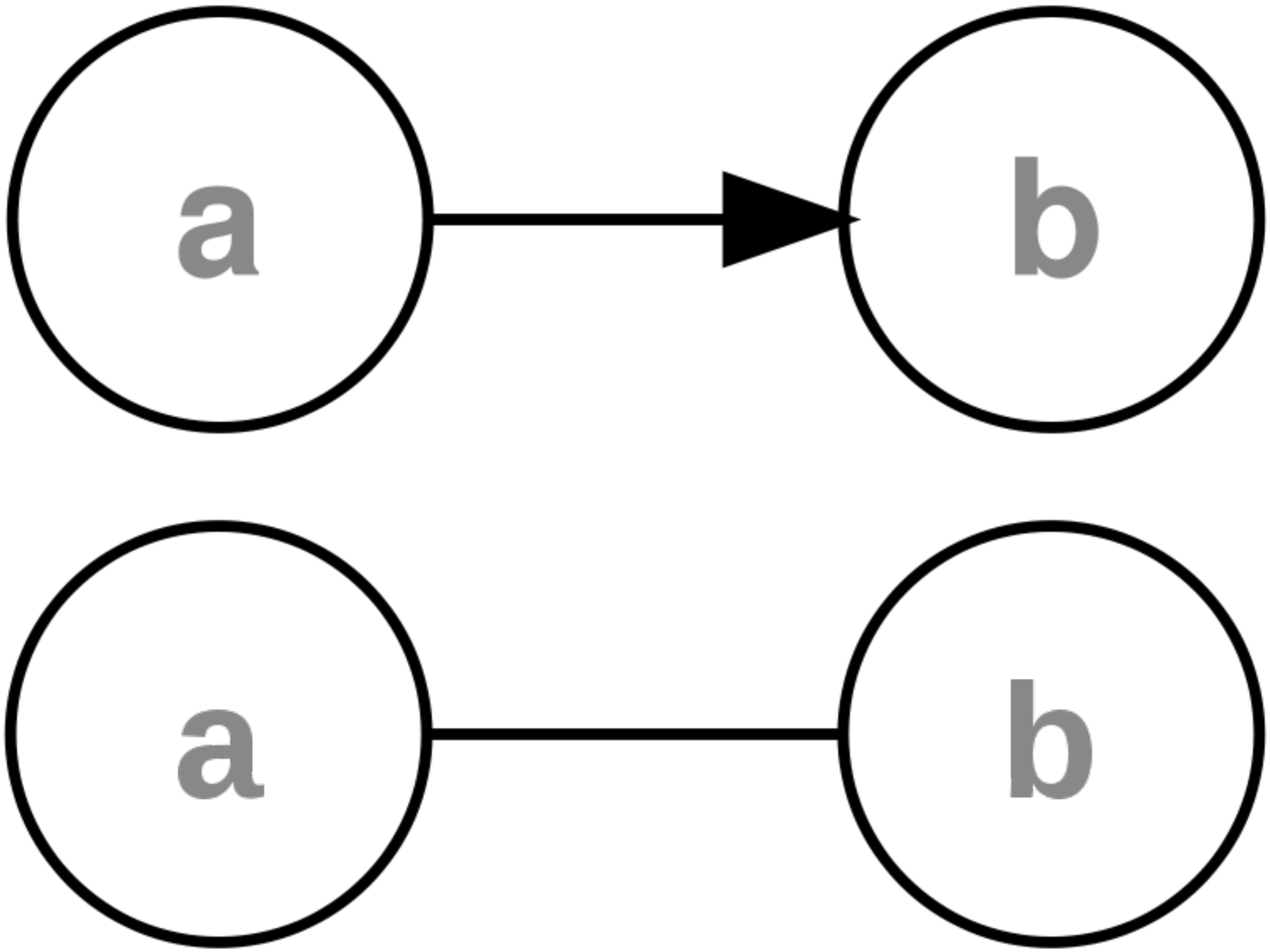
Directed Graph



Undirected Graph



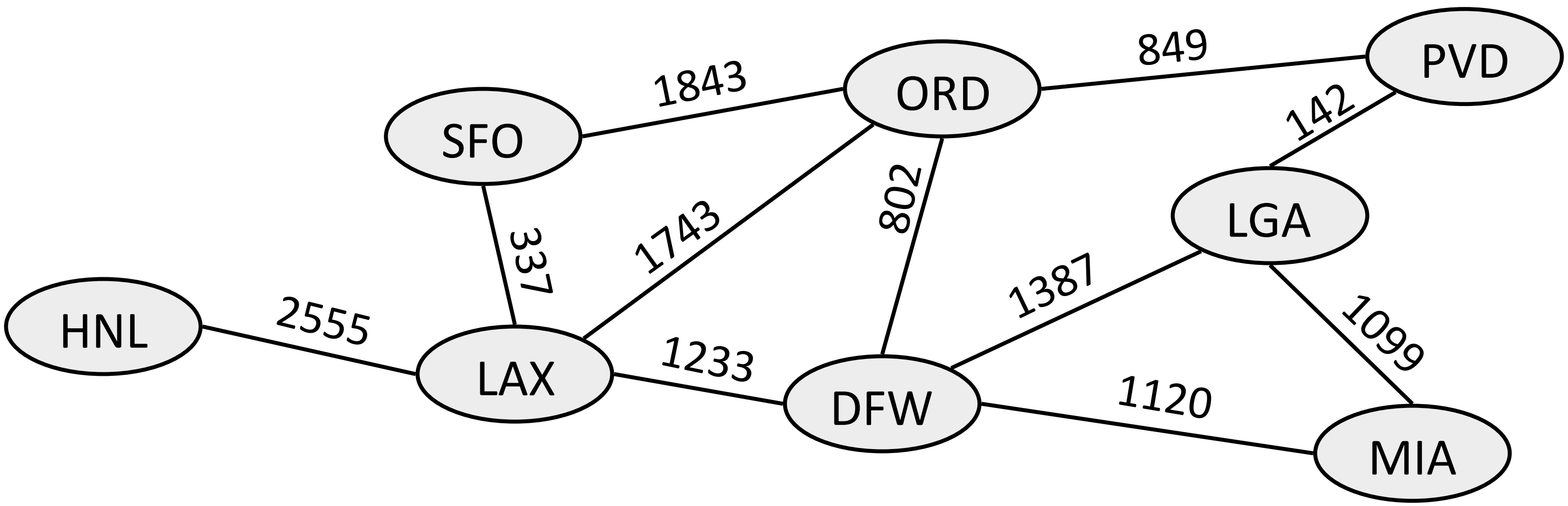
Directed vs Undirected



Weighted graphs

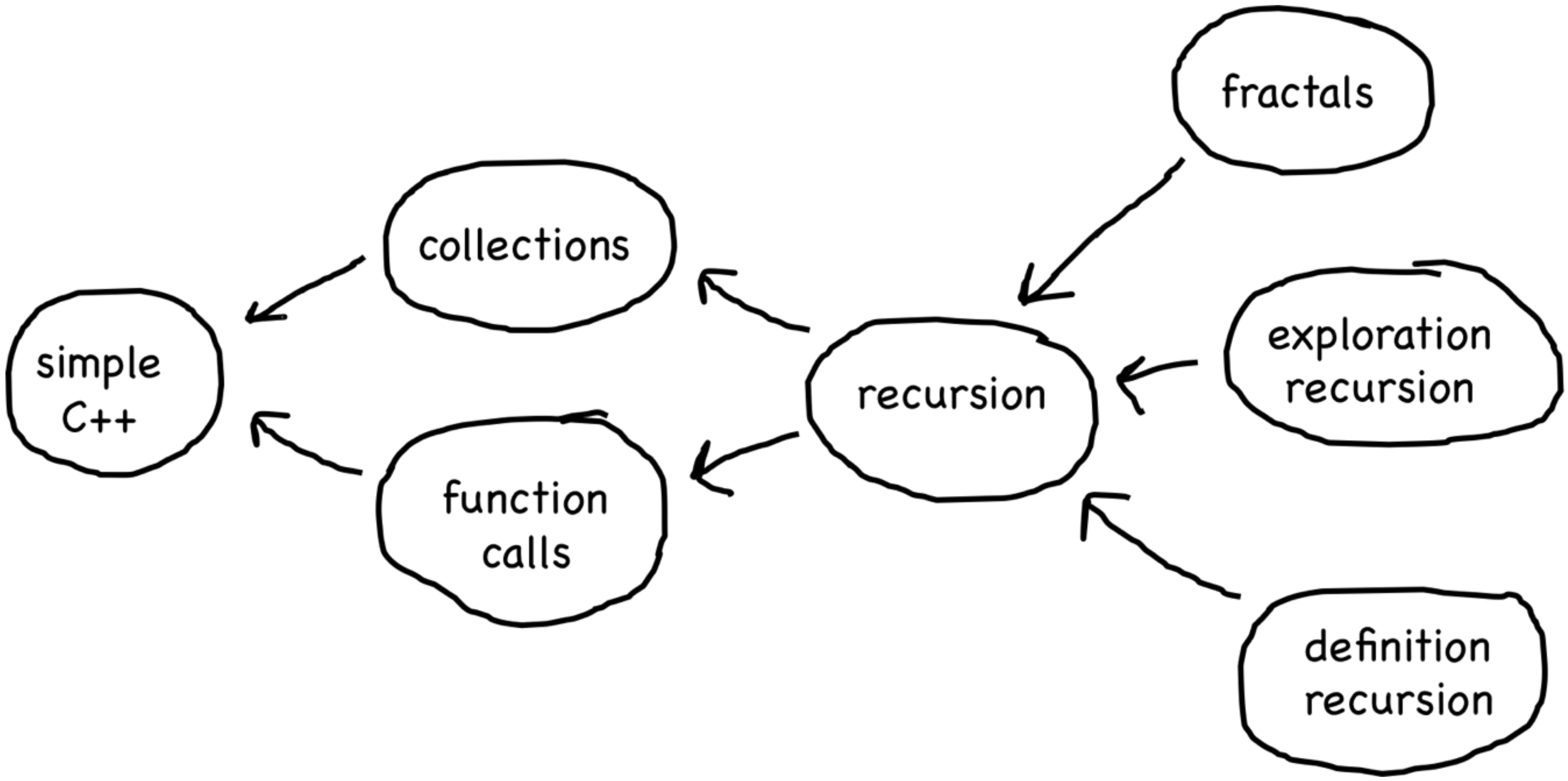
weight: Cost associated with a given edge.

example: graph of airline flights, weighted by miles between cities:

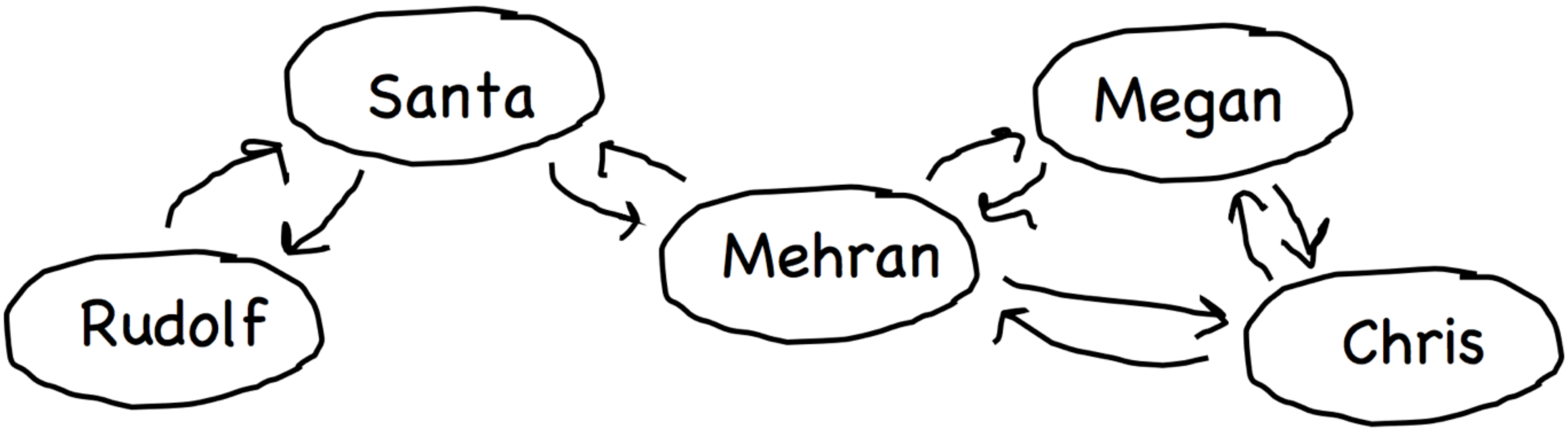


You have seen this before…

Prerequisite Graph

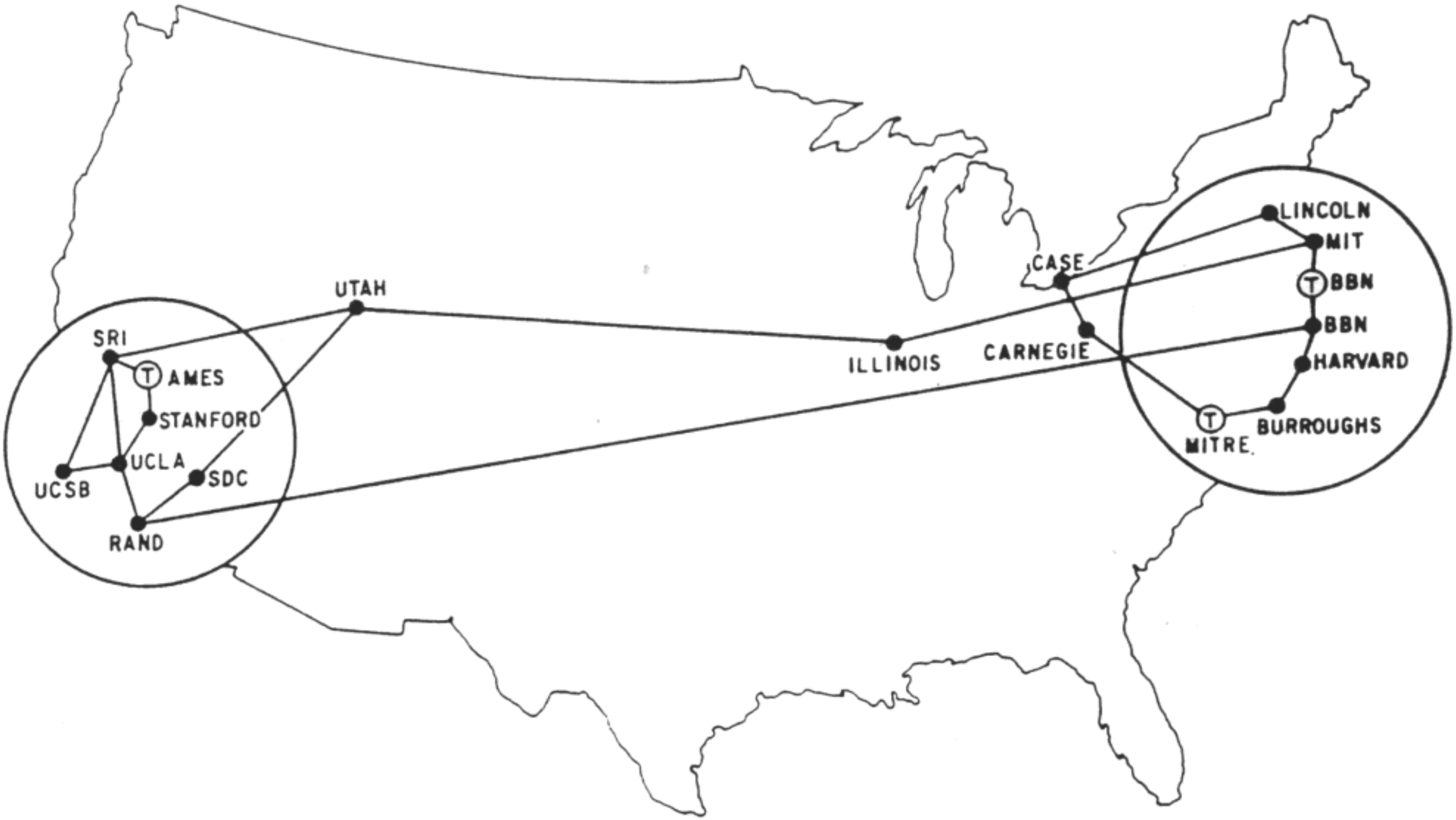


Social Network Graph



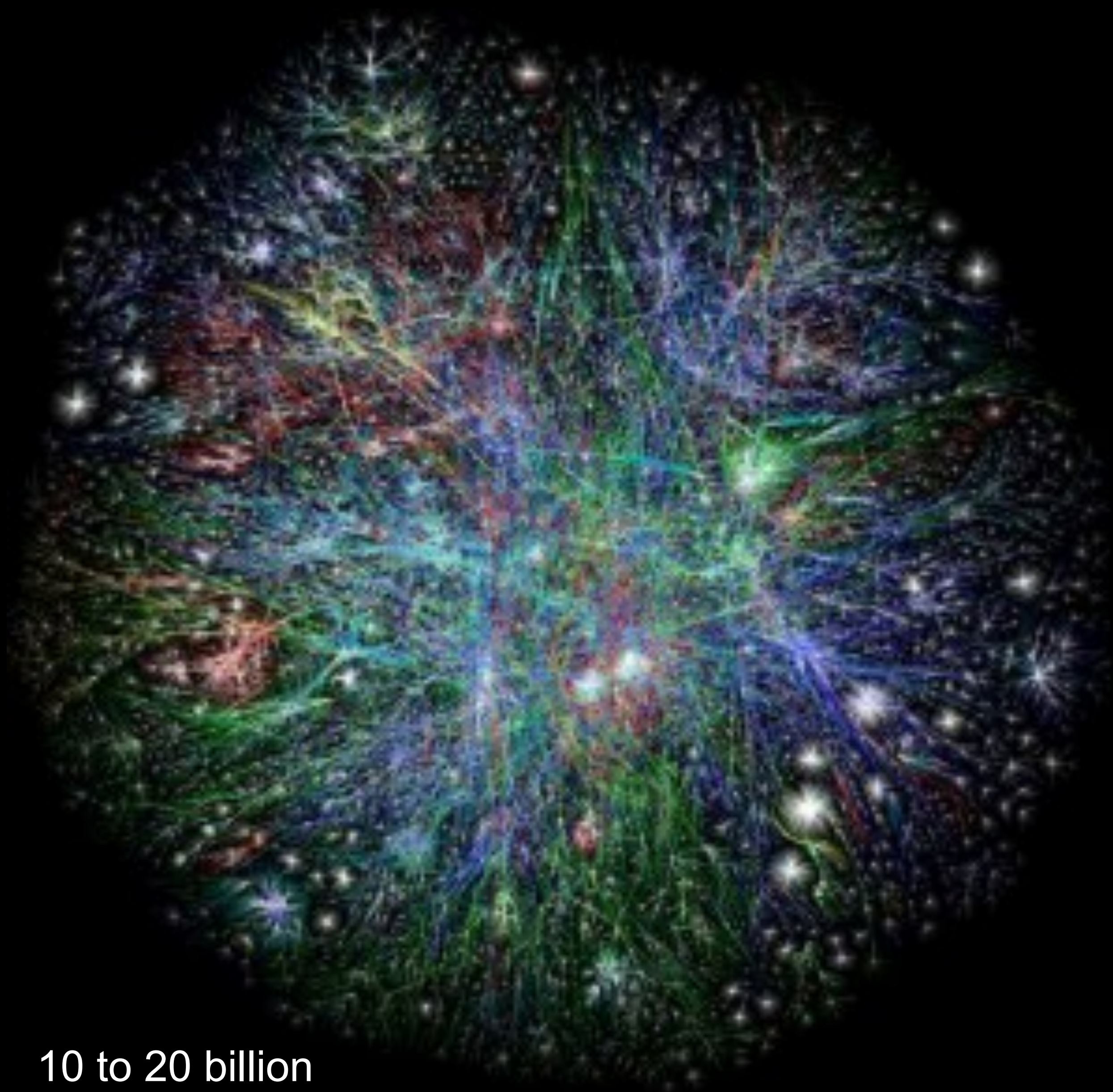
Social Network

The Internet



MAP 4 September 1971

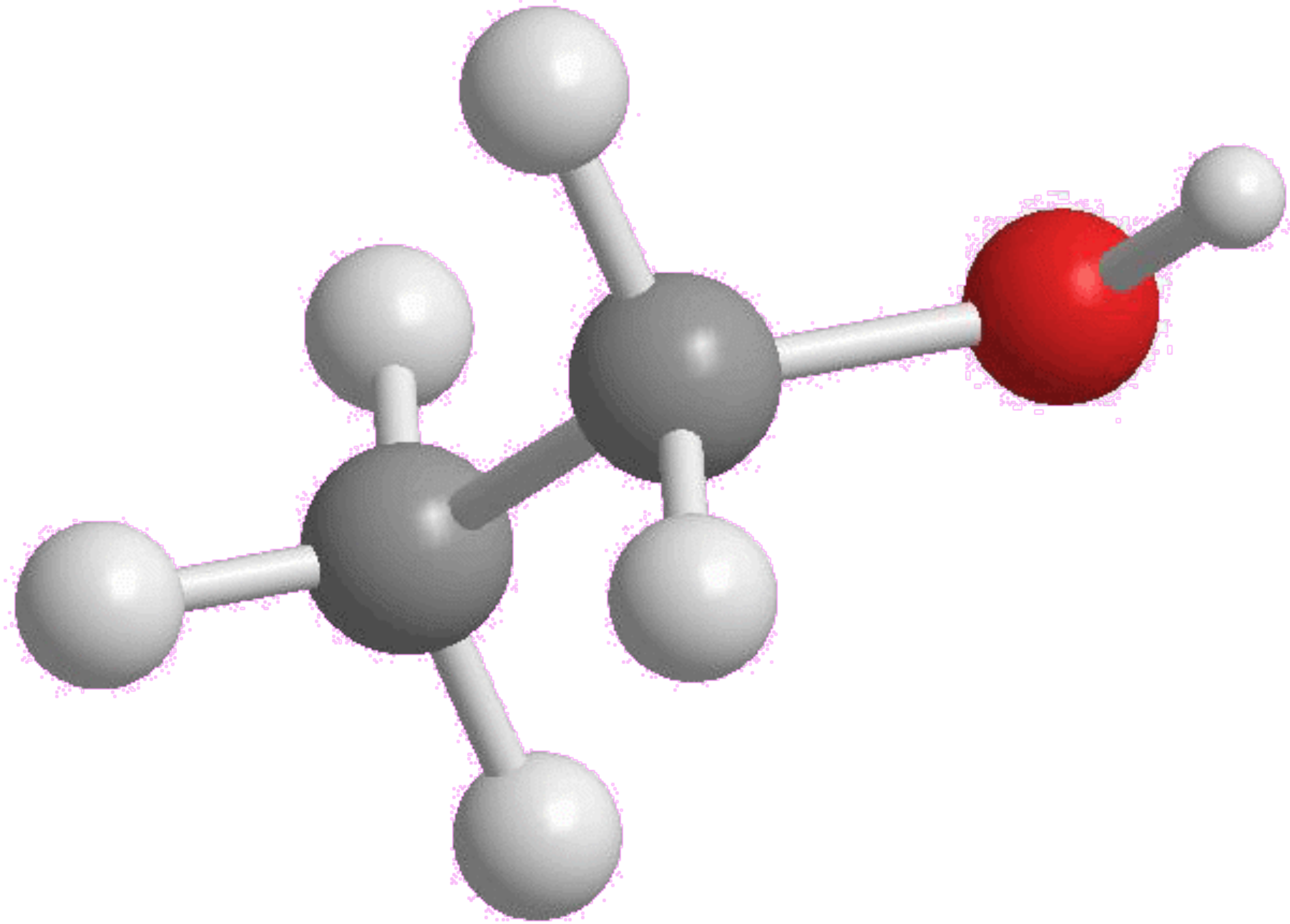
The Internet



10 to 20 billion

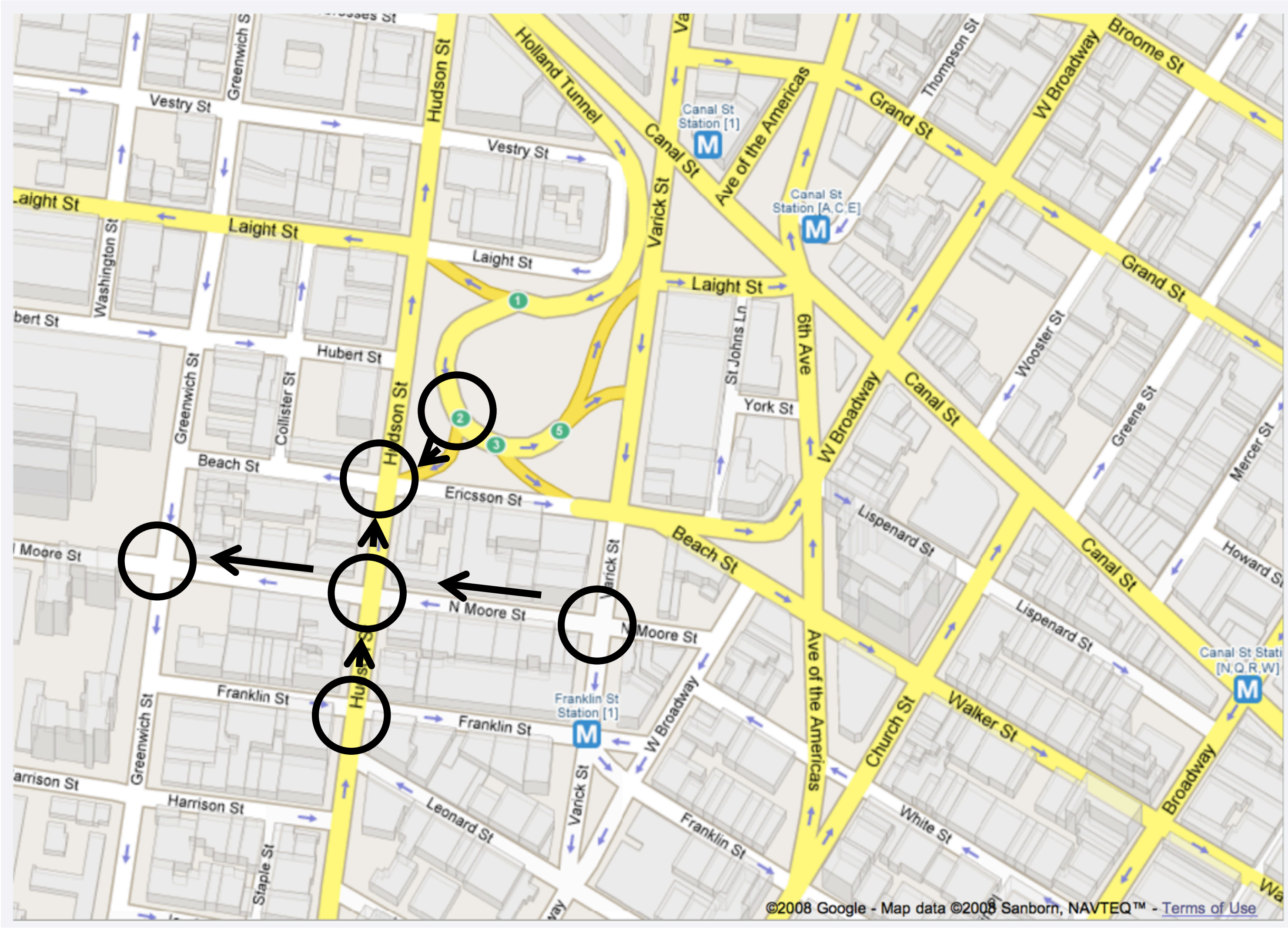
CS Assignments

Chemical Bonds

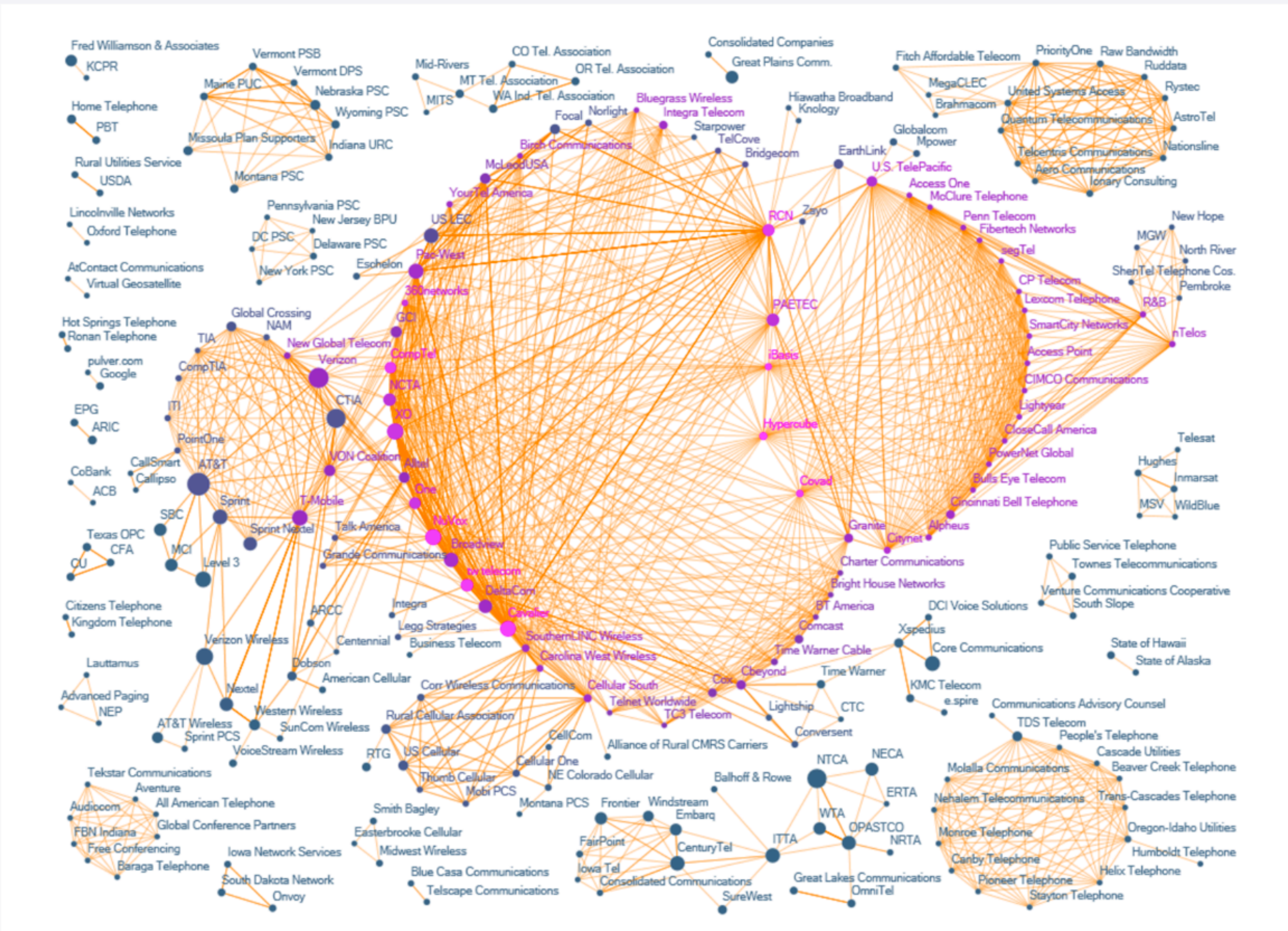


http://4.bp.blogspot.com/-xCtBJ8lKHqA/Tjm0BONWBRI/AAAAAAAAAK4/-mHrbAUOHHg/s1600/Ethanol2.gif

Road Map

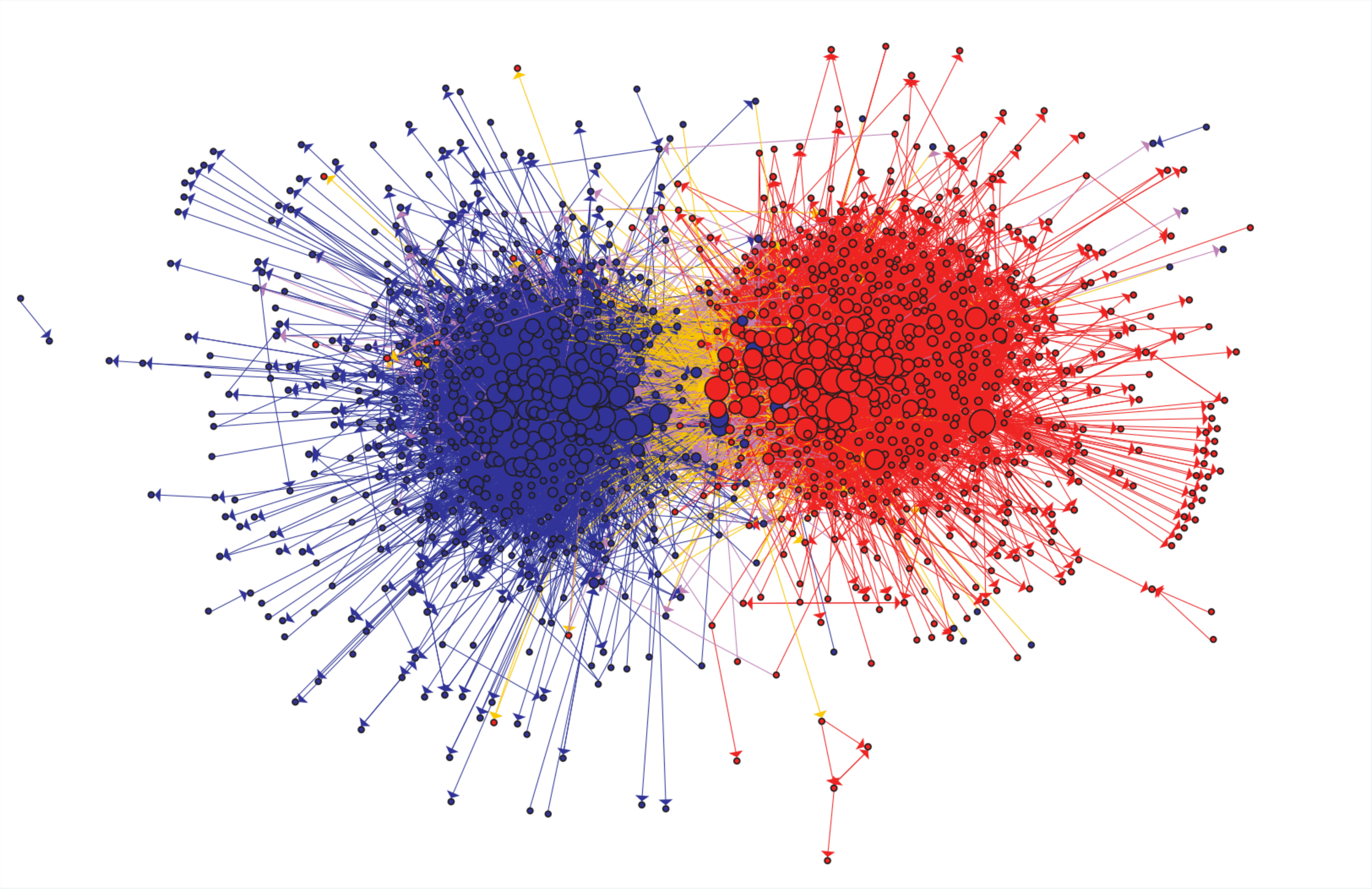

Vestry St
Laight St
Hudson St
Holland Tunnel
Canal St
Varick St
Ave of the Americas
Grand St
Thompson St
W Broadway
Broome St
Canal St Station [1]
Canal St Station [A,C,E]
Washington St
Greenwich St
Hubert St
Collister St
St Johns Ln
York St
6th Ave
Wooster St
Greene St
Mercer St
Beach St
Ericsson St
Lispenard St
Howard St
Moore St
N Moore St
Franklin St
Franklin St Station [1]
Canal St Station [N,Q,R,W]
Walker St
Church St
Harrison St
Leonard St
White St
Broadway
Staple St
©2008 Google - Map data ©2008 Sanborn, NAVTEQ™ - Terms of Use

Corruption



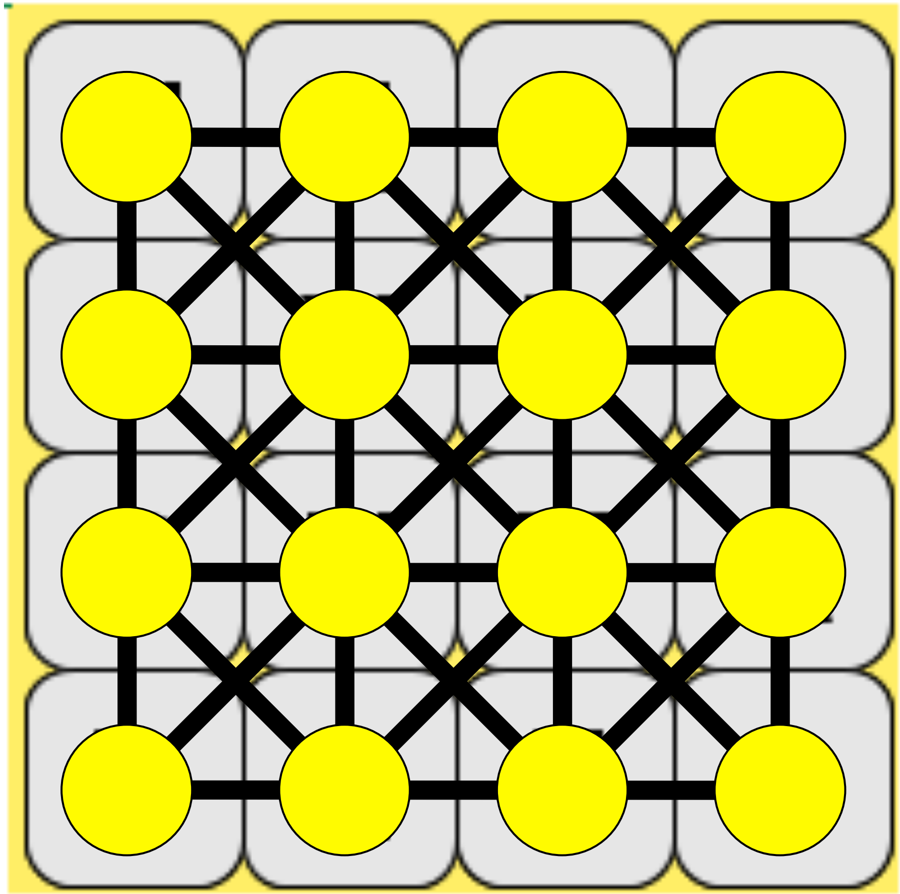
"The Evolution of FCC Lobbying Coalitions" by Pierre de Vries in JoSS Visualization Symposium 2010

Partisanship

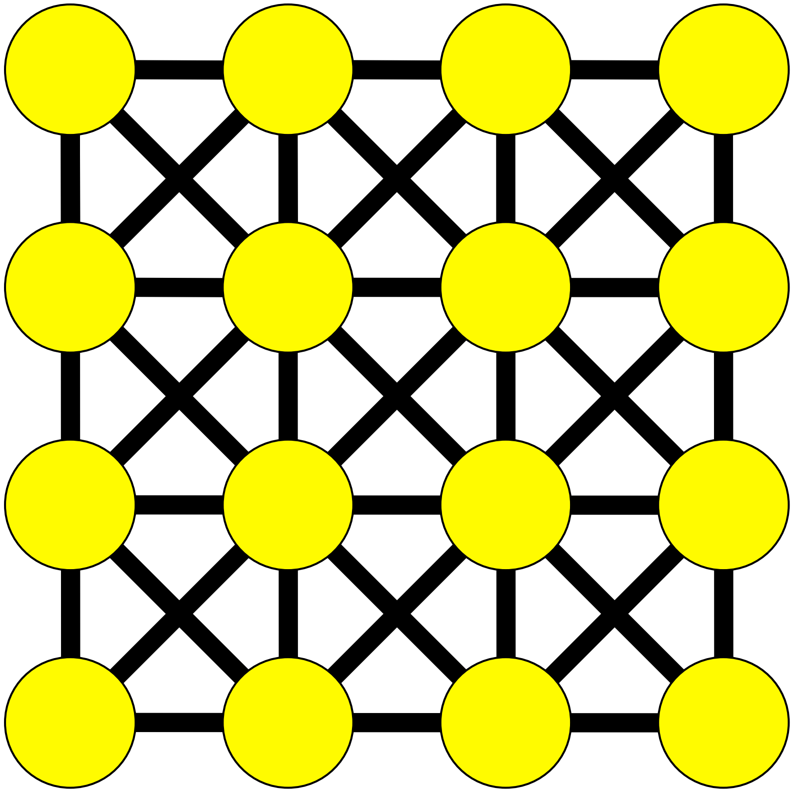


Boggle

Boggle



Boggle



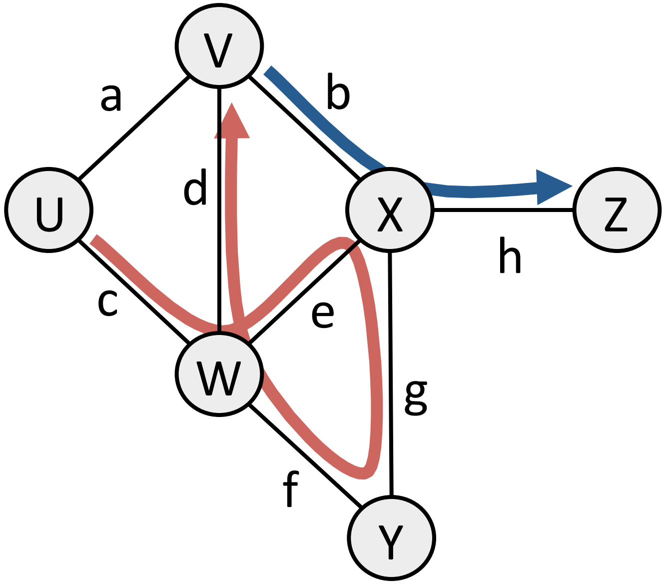
Some terms:

Paths

- **path**: A path from vertex *a* to *b* is a sequence of edges that can be followed starting from *a* to reach *b*.
  - can be represented as vertices visited, or edges taken
  - example, one path from *V* to *Z*: {b, h} or {V, X, Z}
  - What are two paths from U to Y?

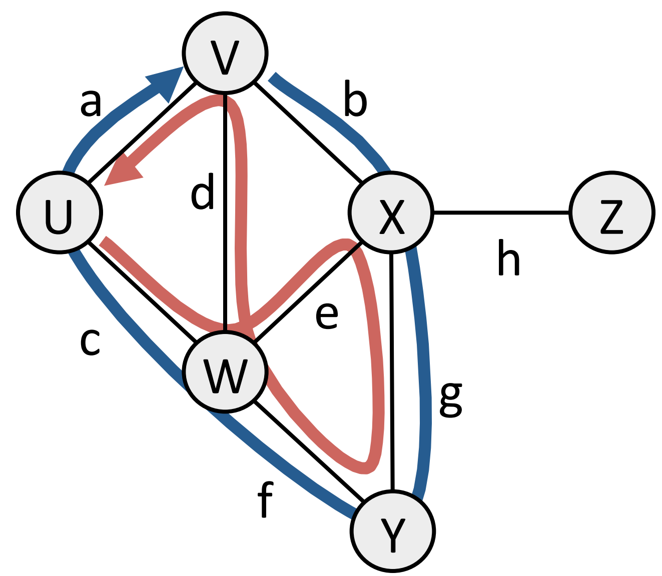
- **path length**: Number of vertices or edges contained in the path.

- **neighbor** or **adjacent**: Two vertices connected directly by an edge.
  - example: V and X



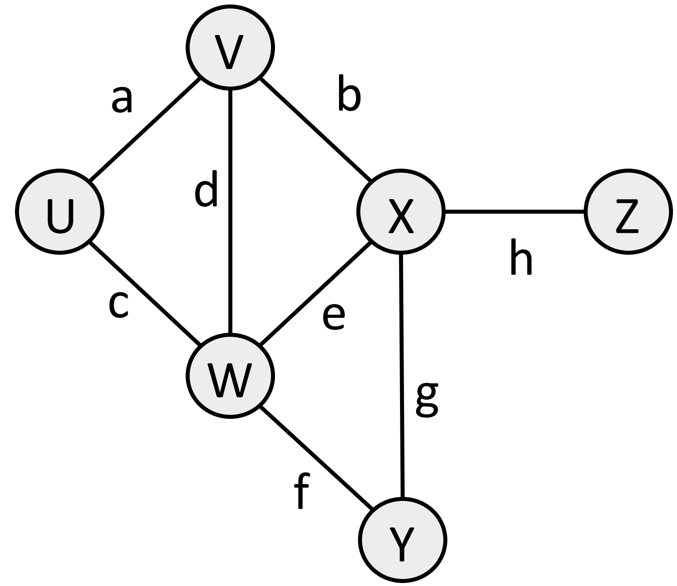
Loops and cycles

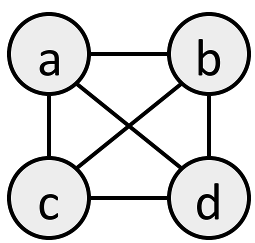
- **cycle**: A path that begins and ends at the same node.
  - example: {b, g, f, c, a} or {V, X, Y, W, U, V}.
  - example: {c, d, a} or {U, W, V, U}.
  - **acyclic graph**: One that does not contain any cycles.
- **loop**: An edge directly from a node to itself.
  - Many graphs don't allow loops.

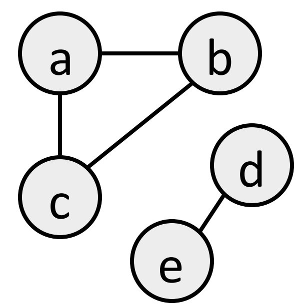


Reachability, connectedness

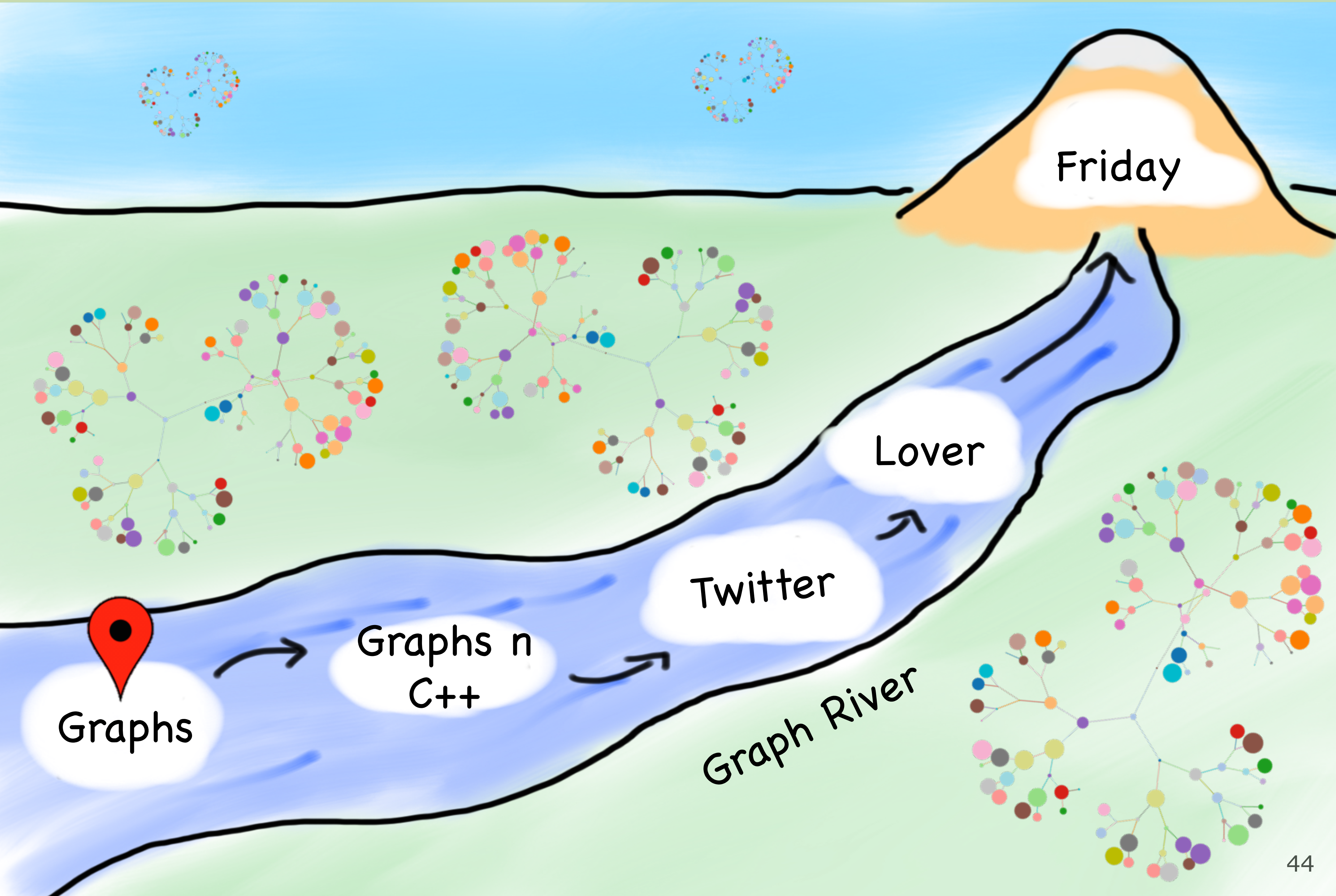
- **reachable**: Vertex *a* is *reachable* from *b* if a path exists from *a* to *b*.
- **connected**: A graph is *connected* if every vertex is reachable from every other.
- **complete**: If every vertex has a direct edge to every other.



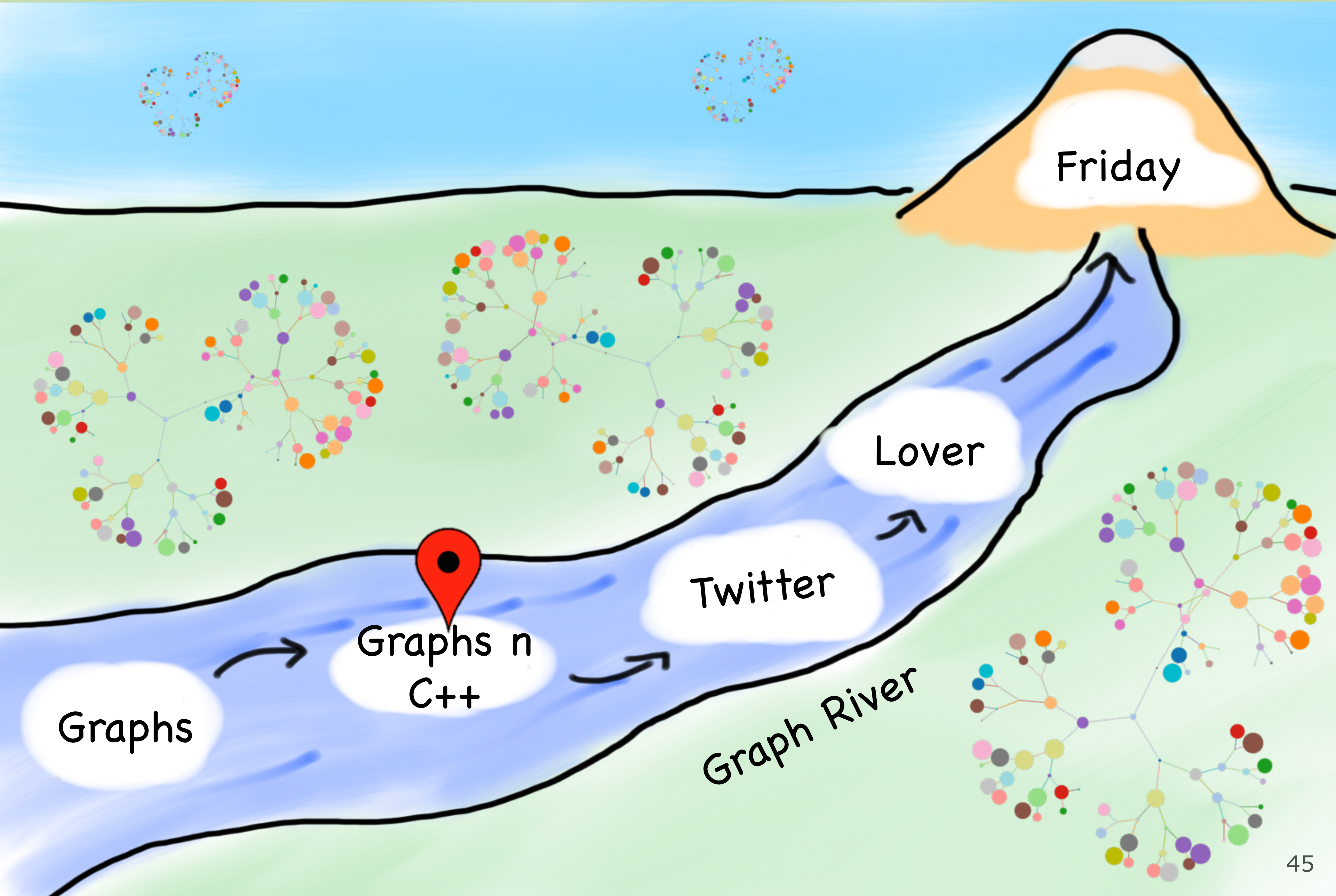




Today's Route



Today's Route



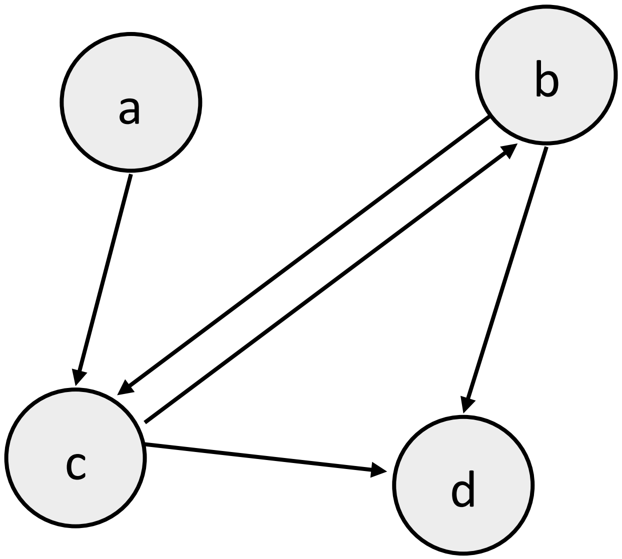
Stanford BasicGraph

The Stanford C++ library includes a `BasicGraph` class.

- Based on an older library class named `Graph`

You can construct a graph and add vertices/edges:

```
#include "basicgraph.h"
...
BasicGraph graph;
graph.addVertex("a");
graph.addVertex("b");
graph.addVertex("c");
graph.addVertex("d");
graph.addEdge("a", "c");
graph.addEdge("b", "c");
graph.addEdge("c", "b");
graph.addEdge("b", "d");
graph.addEdge("c", "d");
```



BasicGraph members

```
#include "basicgraph.h"    // a directed, weighted graph
```

g`.addEdge(`***v1***`, `***v2***`);`	adds an edge between two vertexes
g`.addVertex(`***name***`);`	adds a vertex to the graph
g`.clear();`	removes all vertexes/edges from the graph
g`.getEdgeSet()` ***g***`.getEdgeSet(`***v***`)`	returns all edges, or all edges that start at ***v***, as a `Set` of pointers
g`.getNeighbors(`***v***`)`	returns a set of all vertices that ***v*** has an edge to
g`.getVertex(`***name***`)`	returns pointer to vertex with the given name
g`.getVertexSet()`	returns a set of all vertexes
g`.isNeighbor(`***v1***`, `***v2***`)`	returns `true` if there is an edge from vertex ***v1*** to ***v2***
g`.isEmpty()`	returns `true` if queue contains no vertexes/edges
g`.removeEdge(`***v1***`, `***v2***`);`	removes an edge from the graph
g`.removeVertex(`***name***`);`	removes a vertex from the graph
g`.size()`	returns the number of vertexes in the graph
g`.toString()`	returns a string such as `"{a, b, c, a -> b}"`

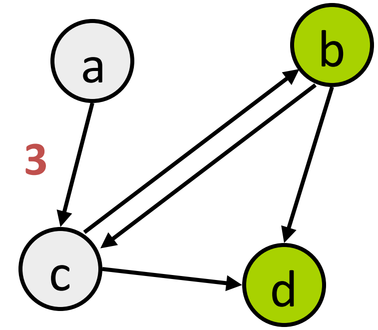
Using BasicGraph

The graph stores a struct of information about each vertex/edge:

```
struct Vertex {
    string name;
    Set<Edge*> edges;
    double cost;
    bool visited;
    Vertex* previous;
};
```

// Non standard

```
struct Edge {
    Vertex* start;
    Vertex* finish;
    double weight;
};
```



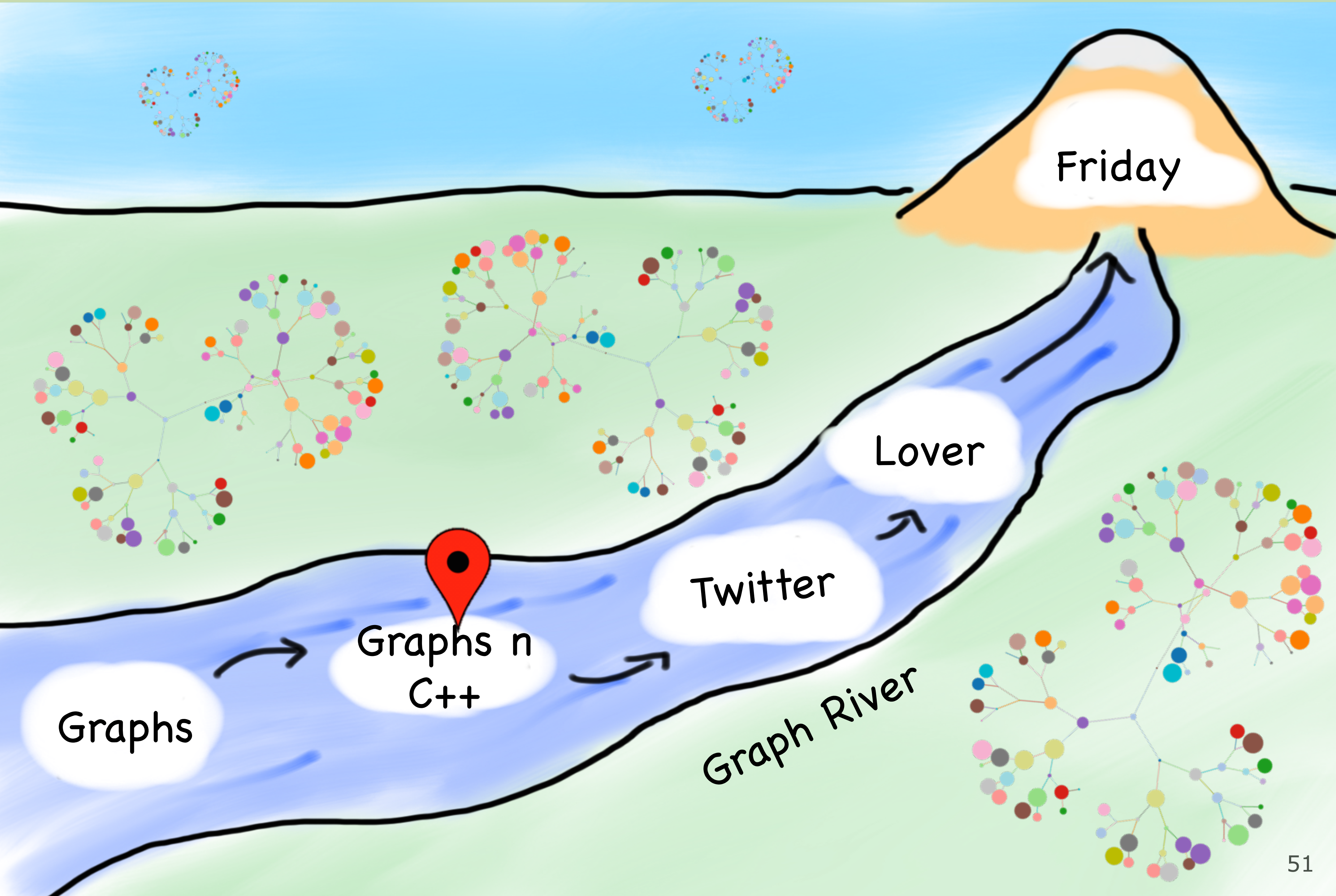
You can use these to help implement graph algorithms:

```
Vertex * vertC = graph.getVertex("c");
Edge * edgeAC = graph.getEdge("a", "c");
```

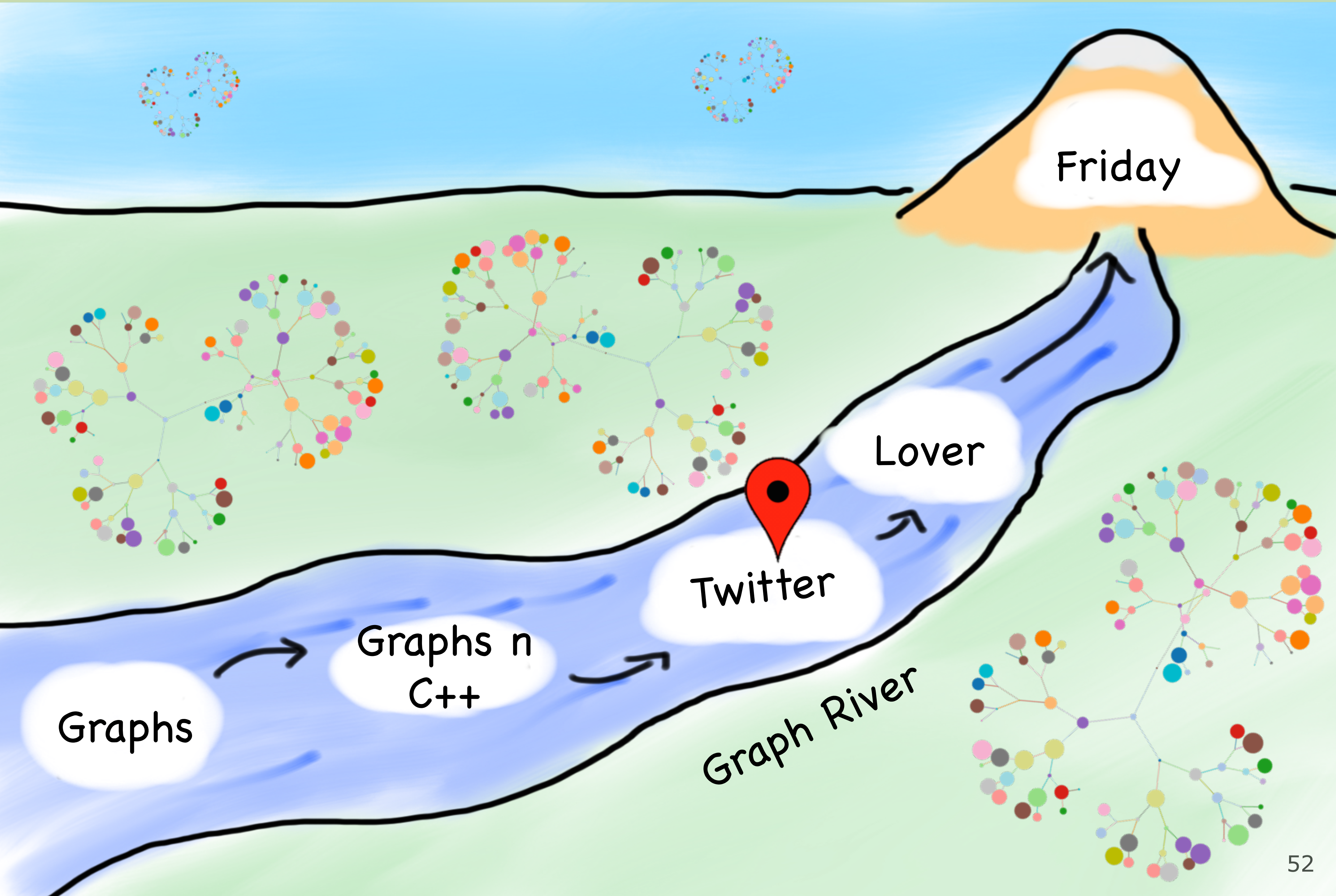
There are other representations…

… this is the one we are going to use.

Today's Route



Today's Route



Algorithms

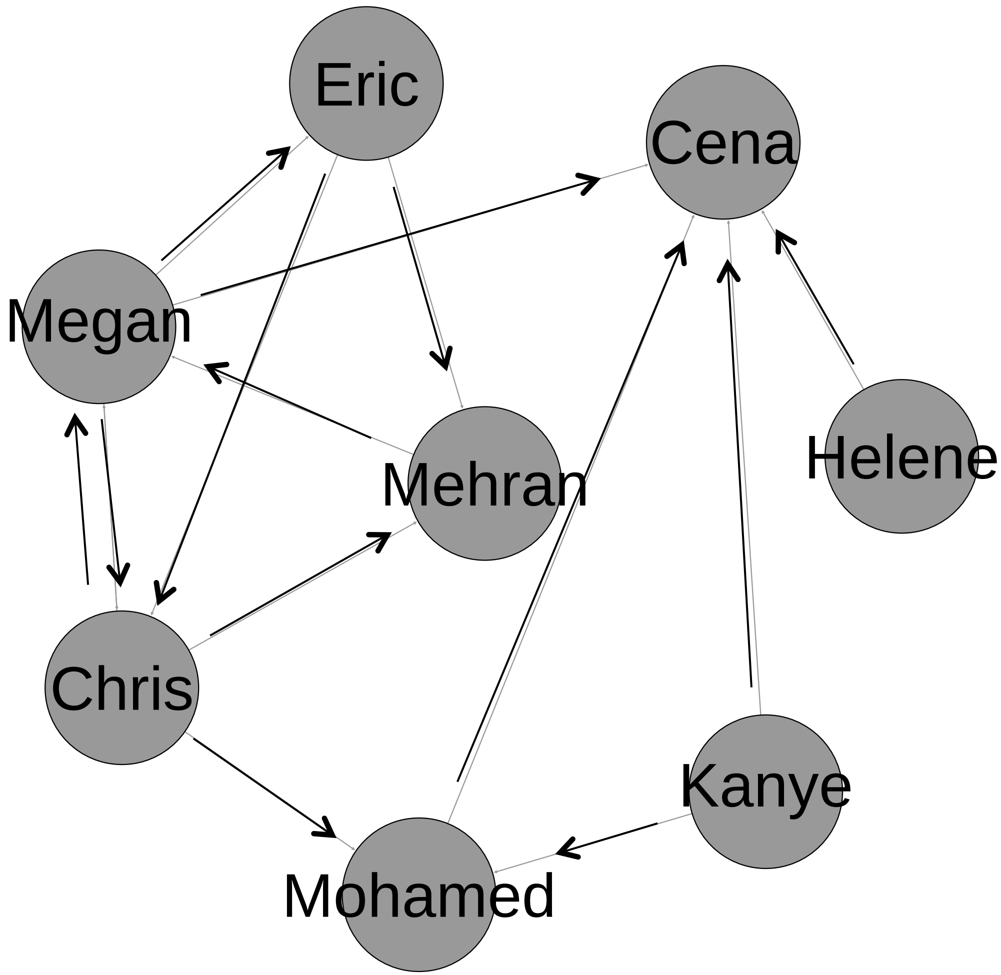
Twitter Popularity

1. Read a file of Twitter followers:

```
id1 id2
id1 id2
```

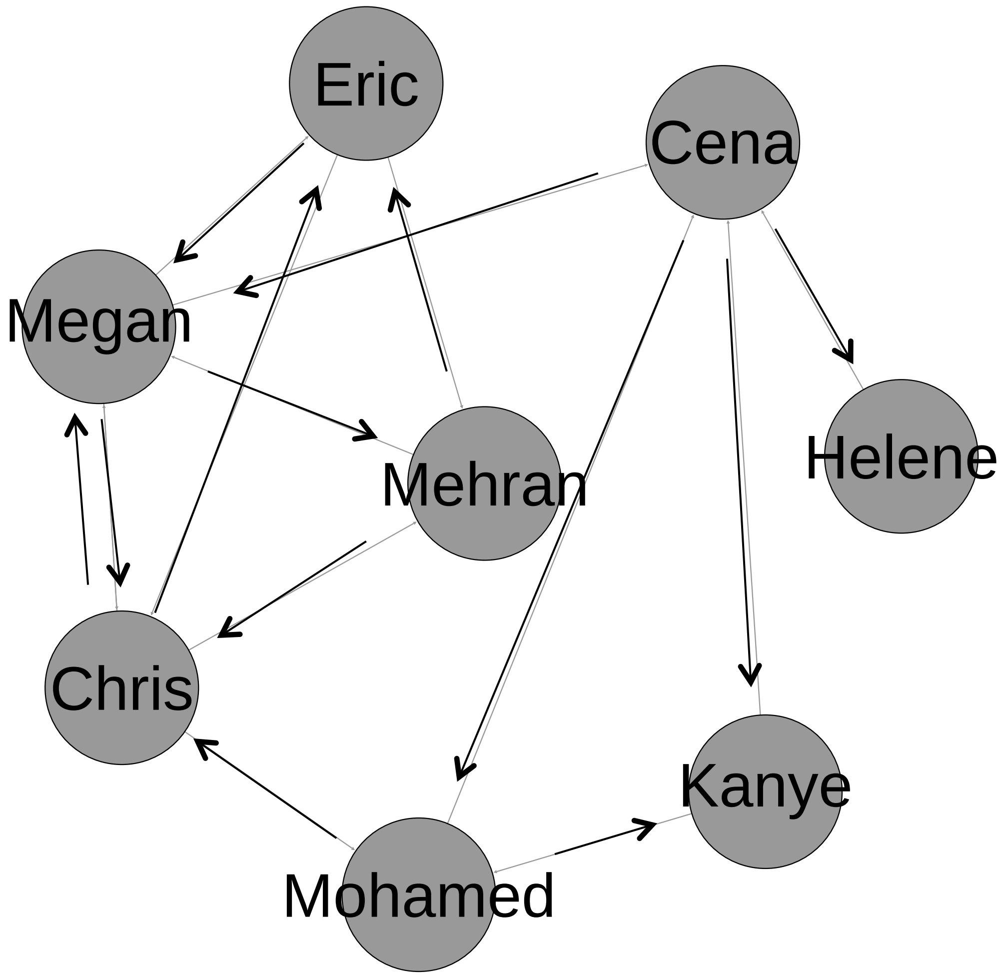
2. Rank users by the number of followers of followers.

BasicGraph exercise



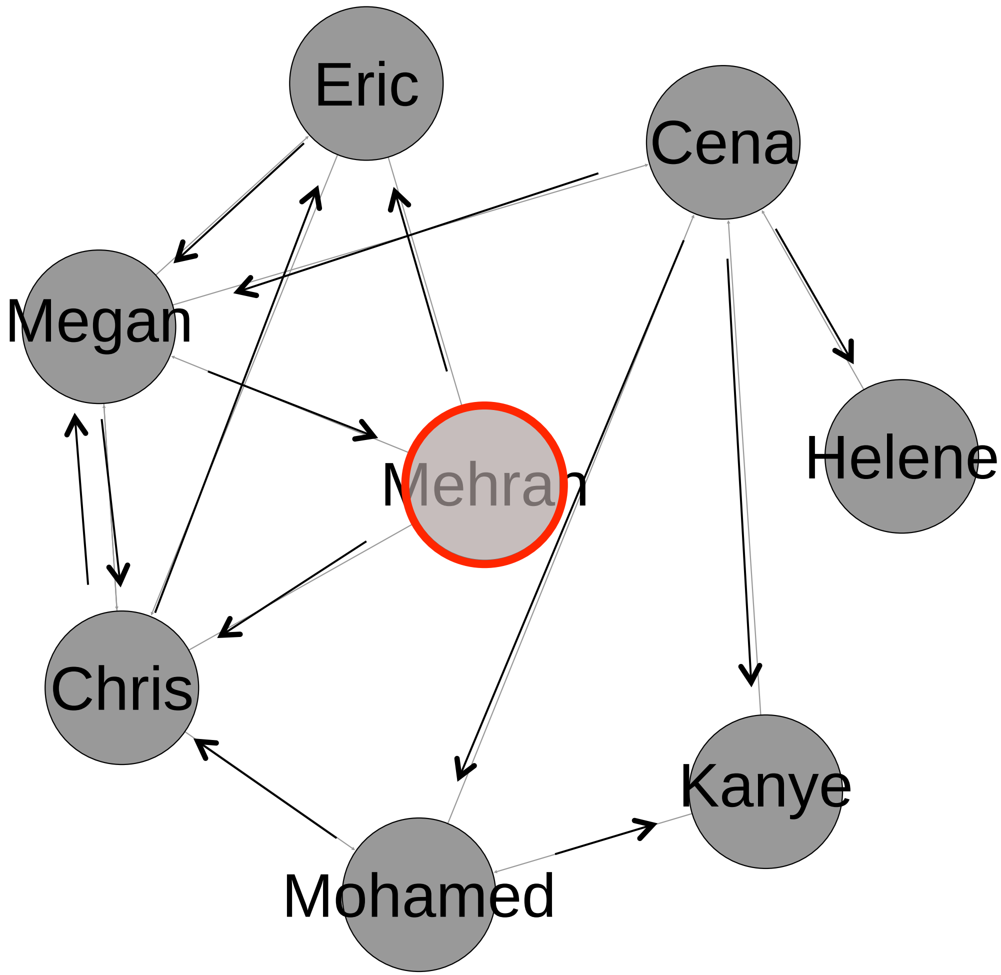
* Start follows finish

BasicGraph exercise



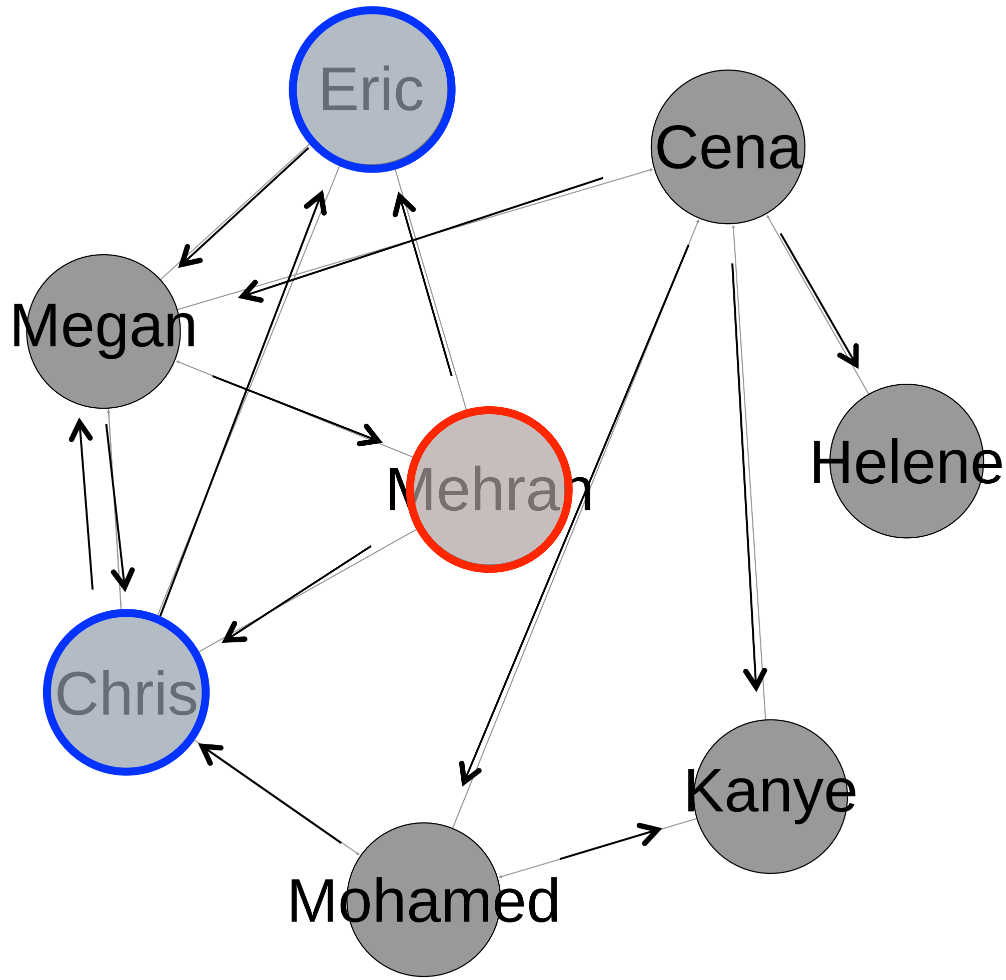
* Finish follows start

BasicGraph exercise



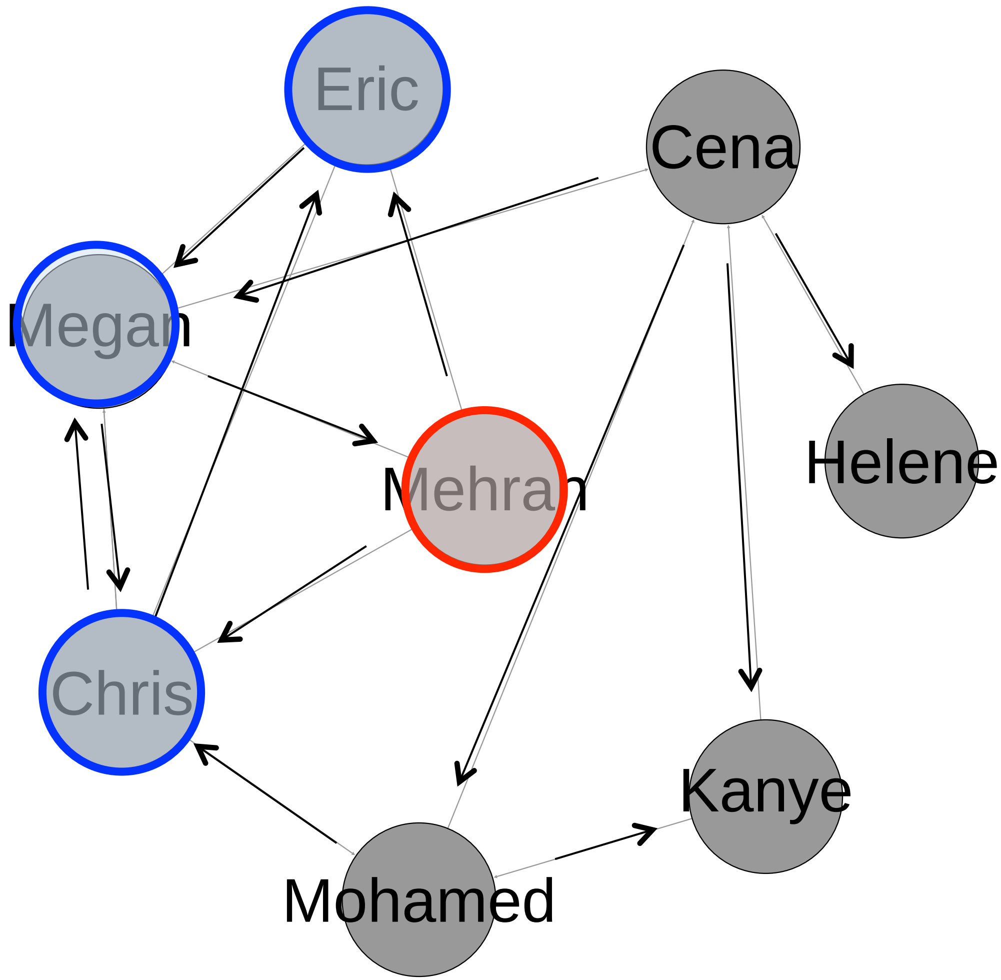
* Finish follows start

First Degree Followers



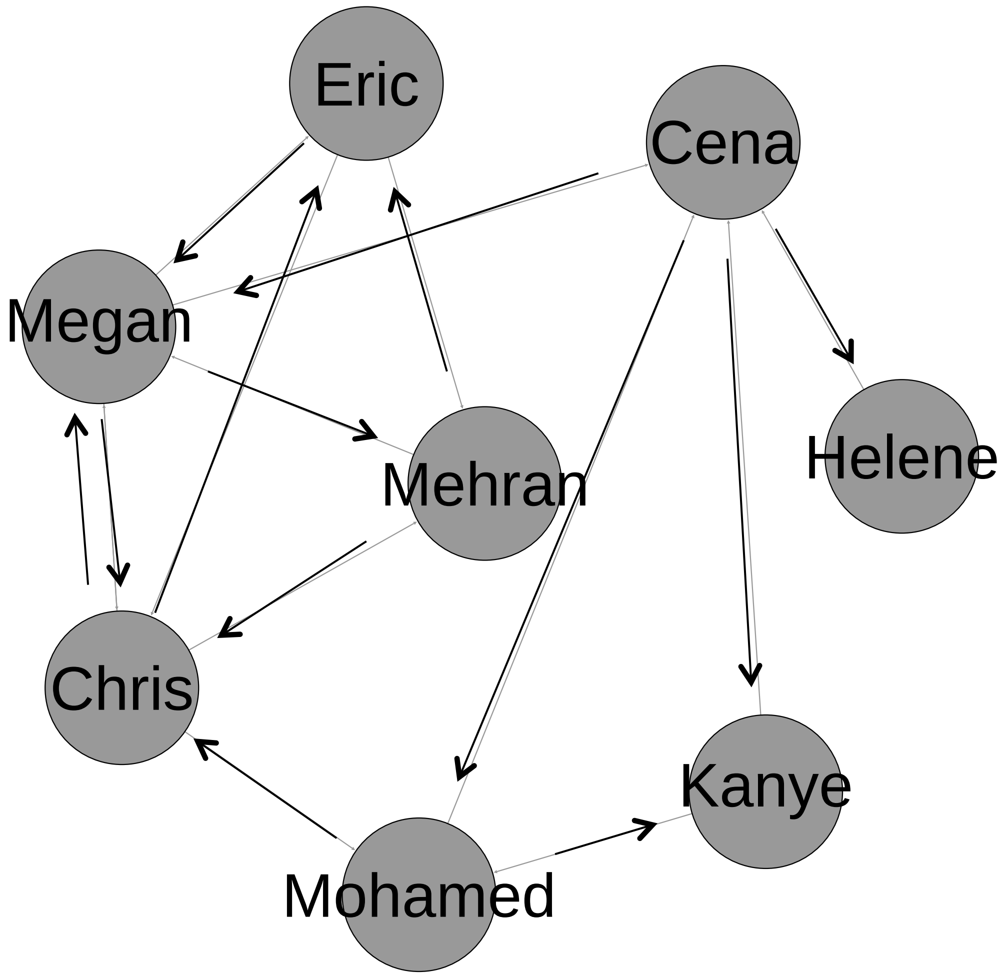
* Finish follows start

Second Degree Followers



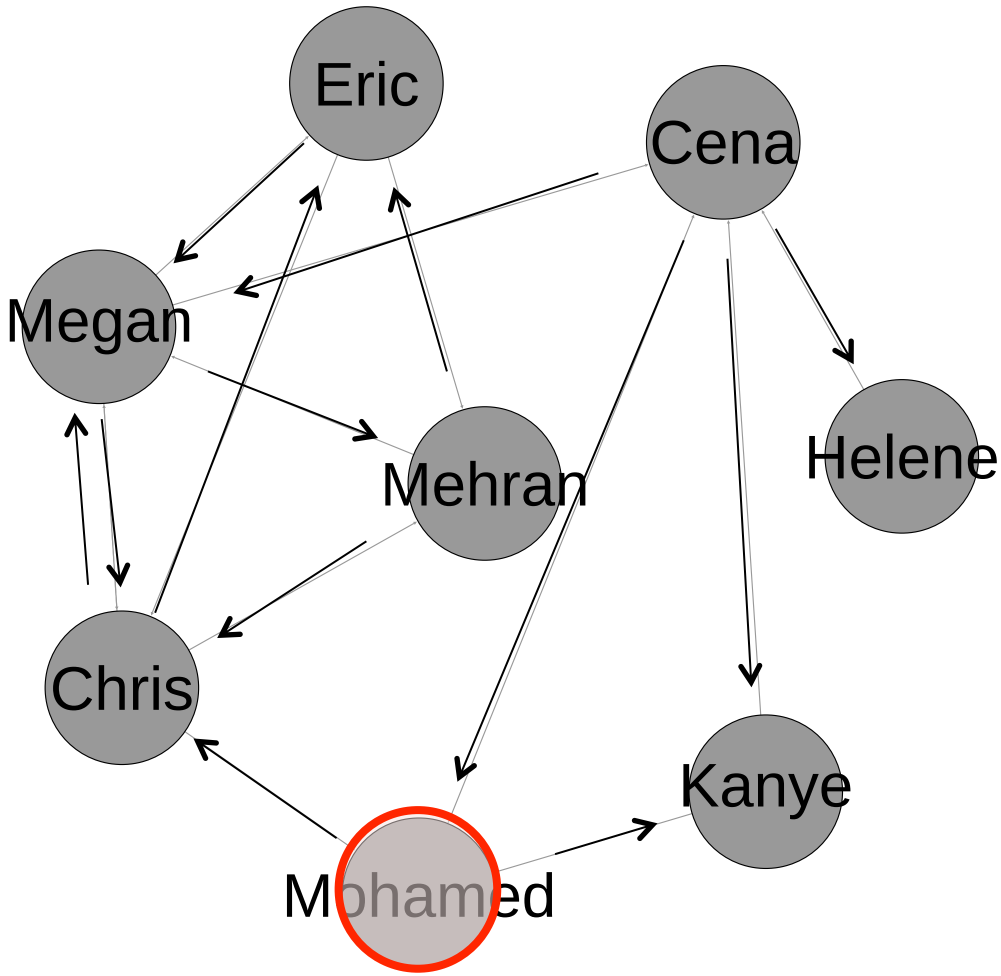
* Finish follows start

BasicGraph exercise



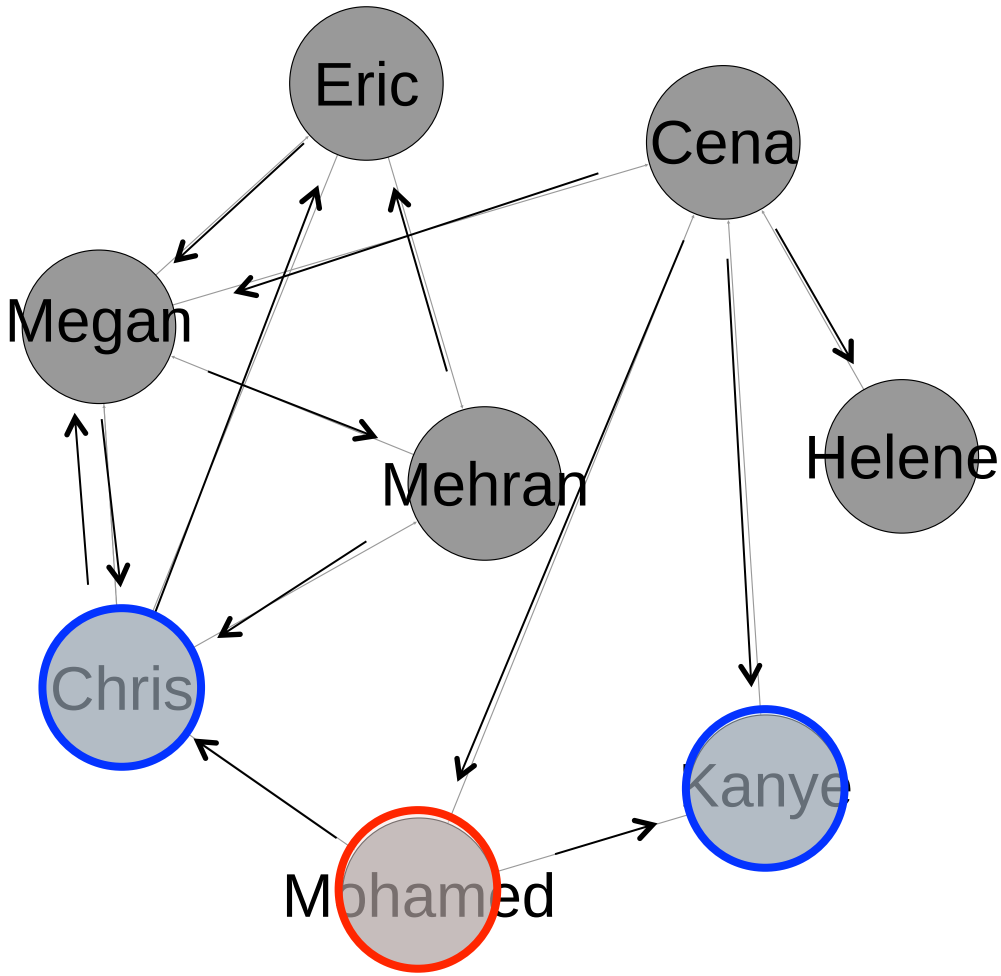
* Finish follows start

BasicGraph exercise



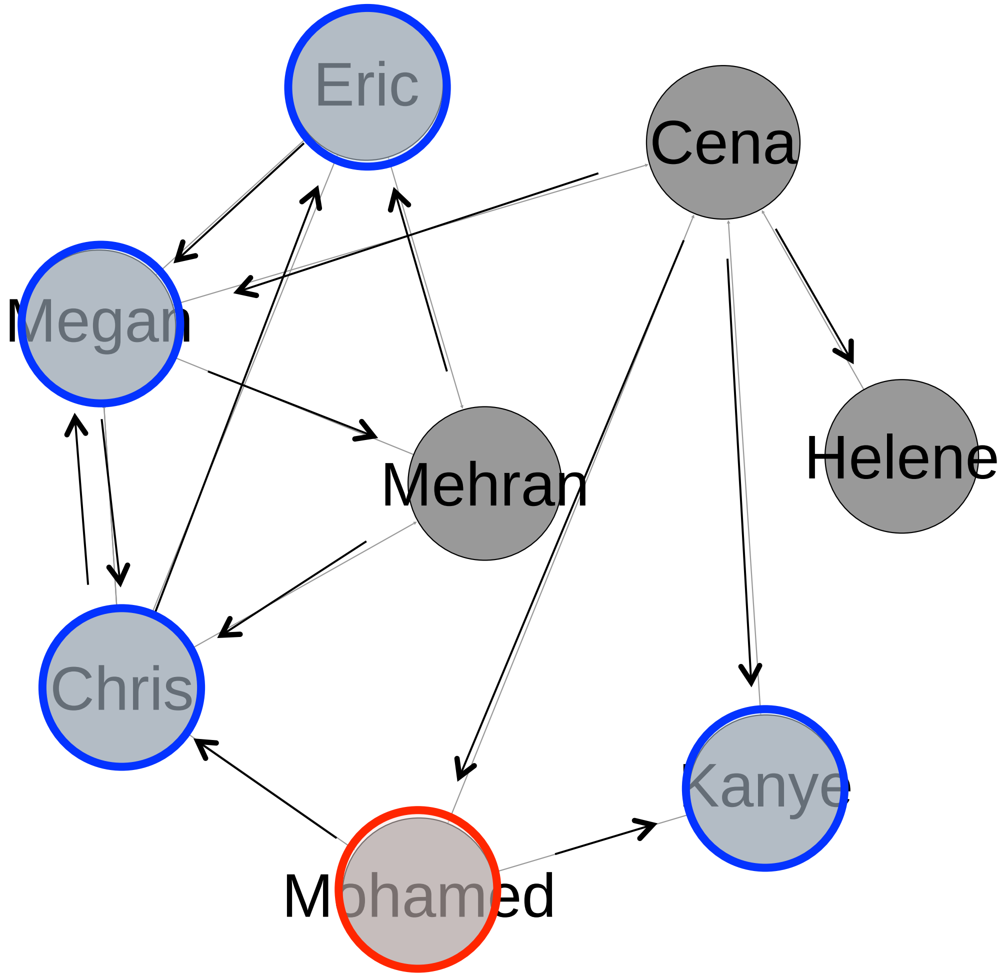
* Finish follows start

First Degree Followers



* Finish follows start

Second Degree Followers



* Finish follows start

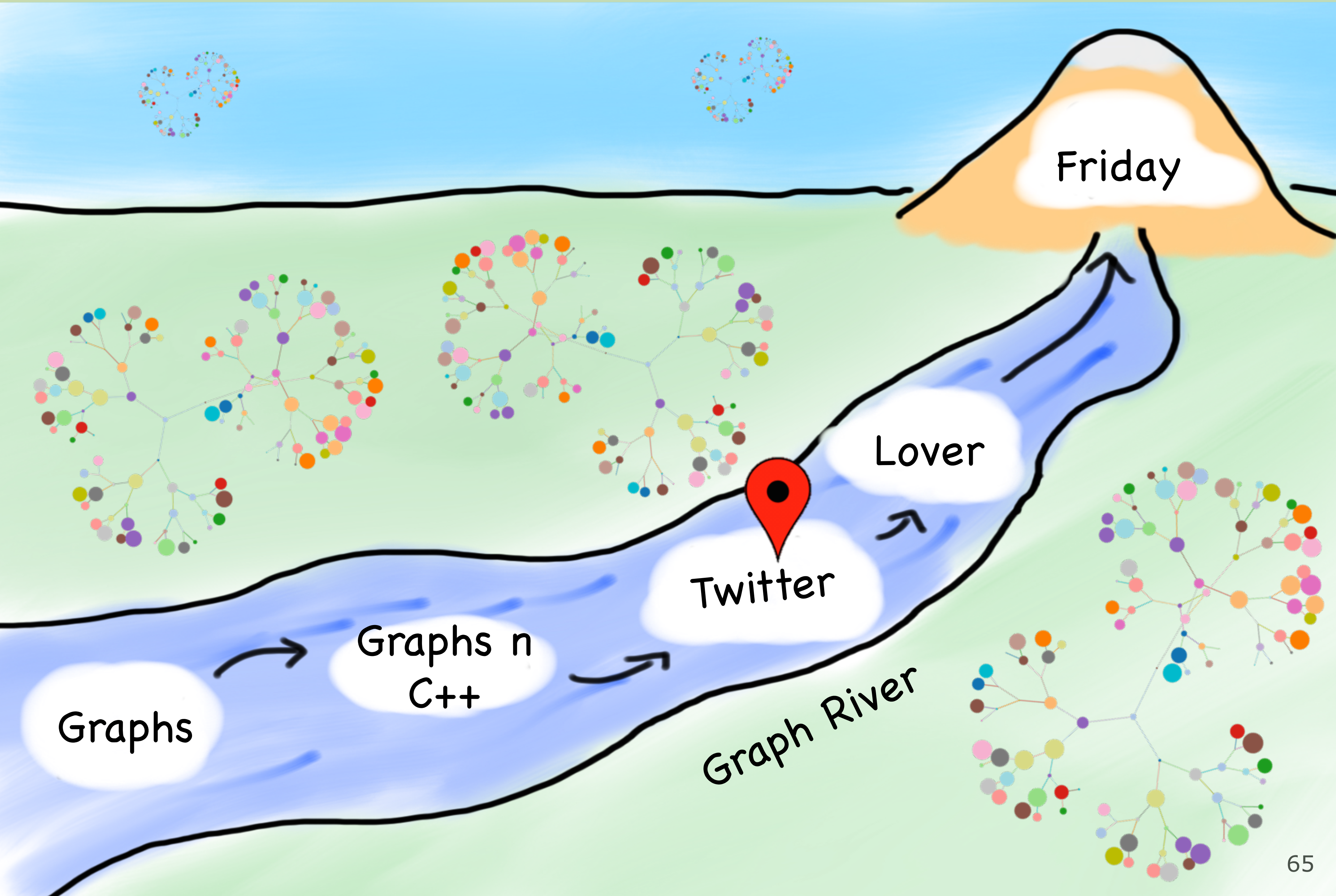
Twitter Popularity

1. Read a file of Twitter followers:

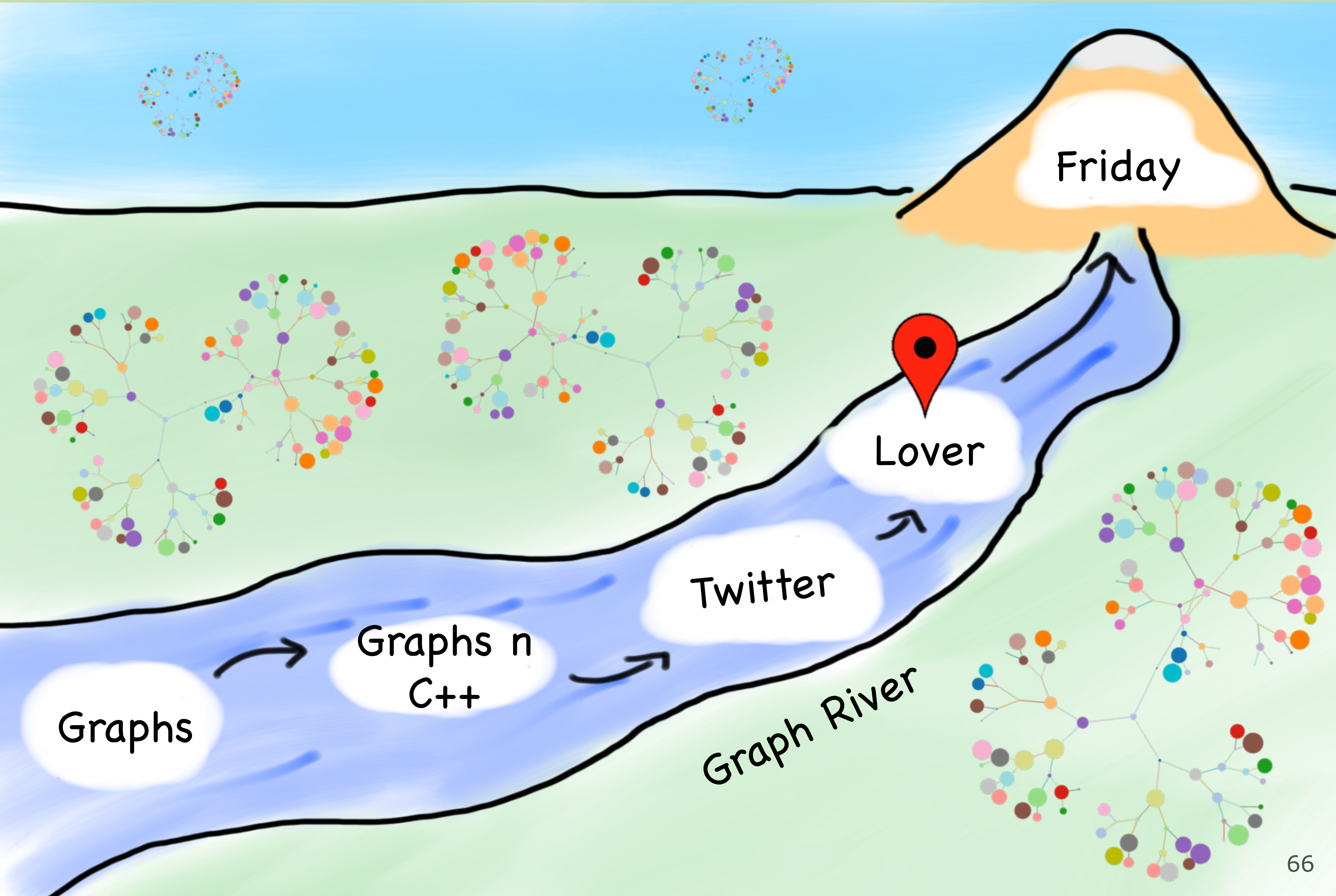
```
id1 id2
id1 id2
```

2. Rank users by the number of followers of followers.

Today's Route

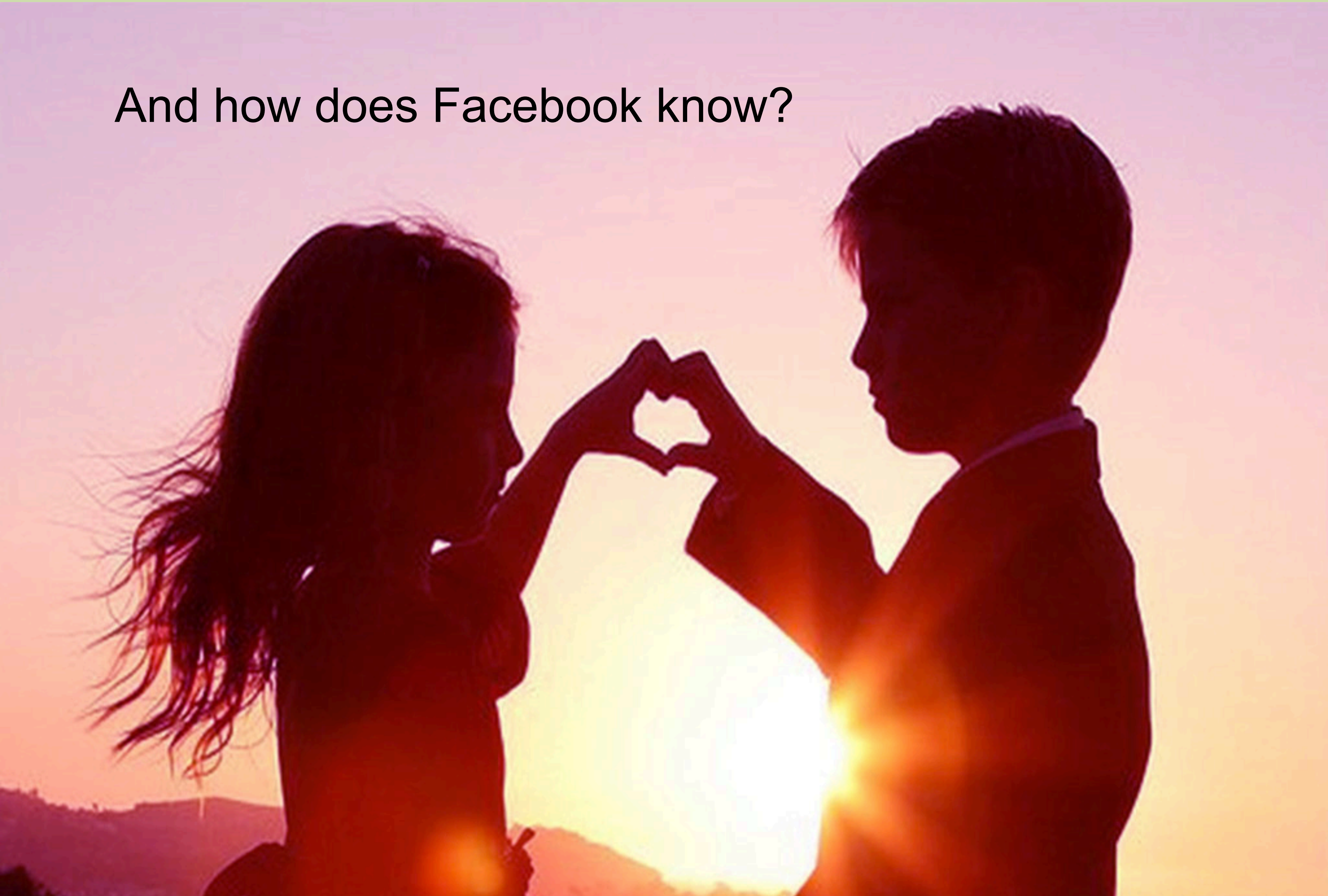


Today's Route

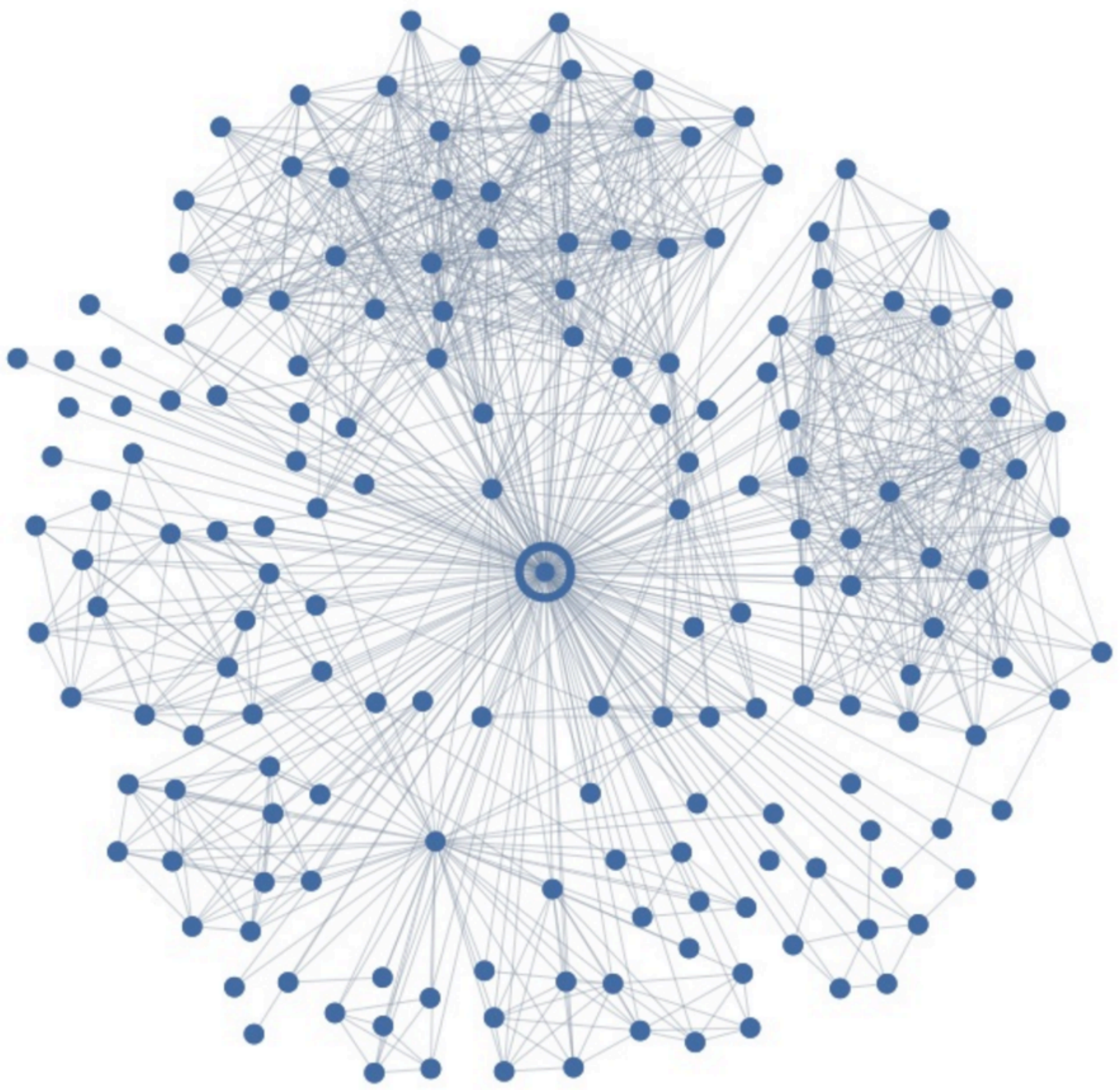


Who Do You Love

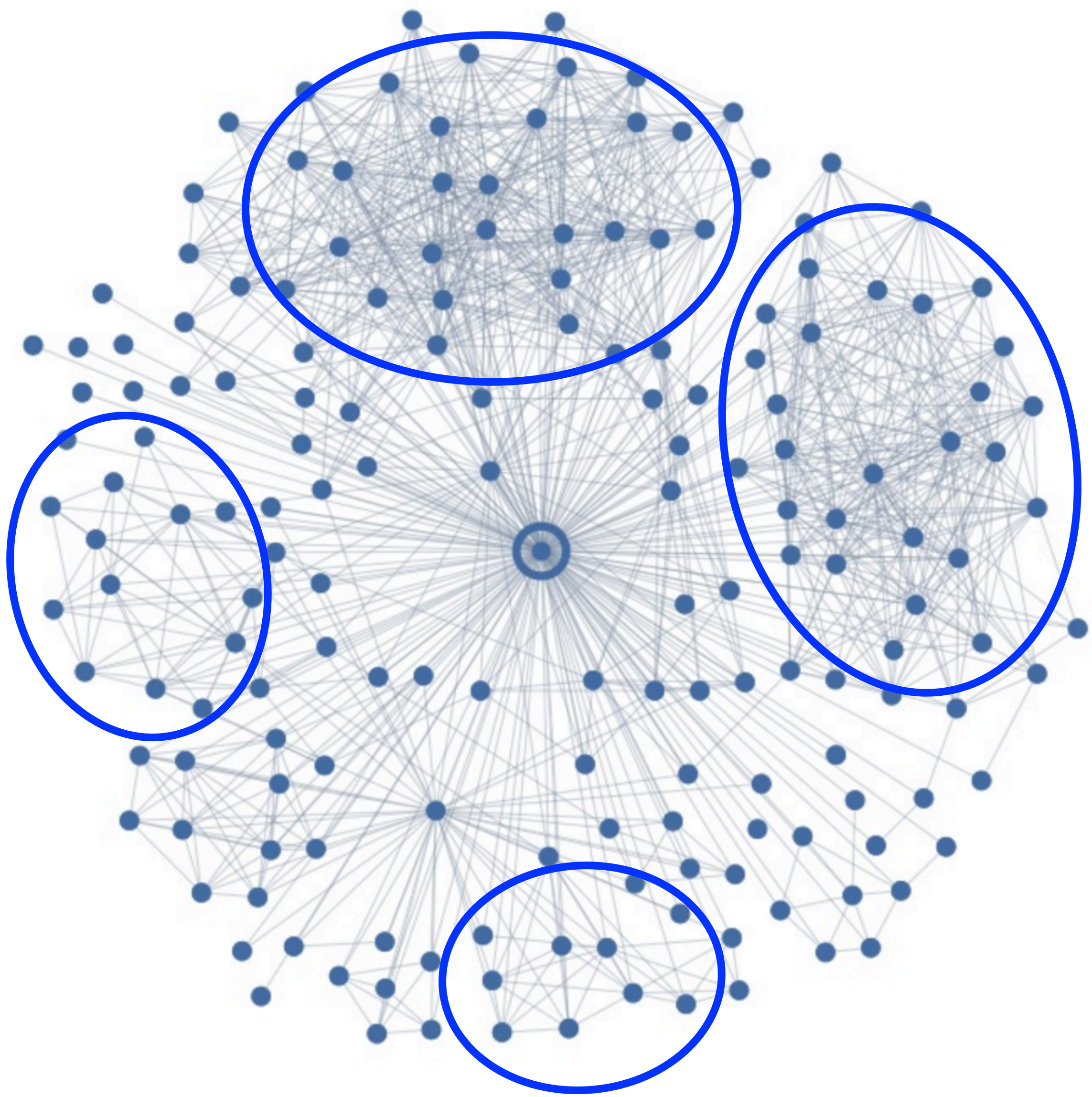
And how does Facebook know?



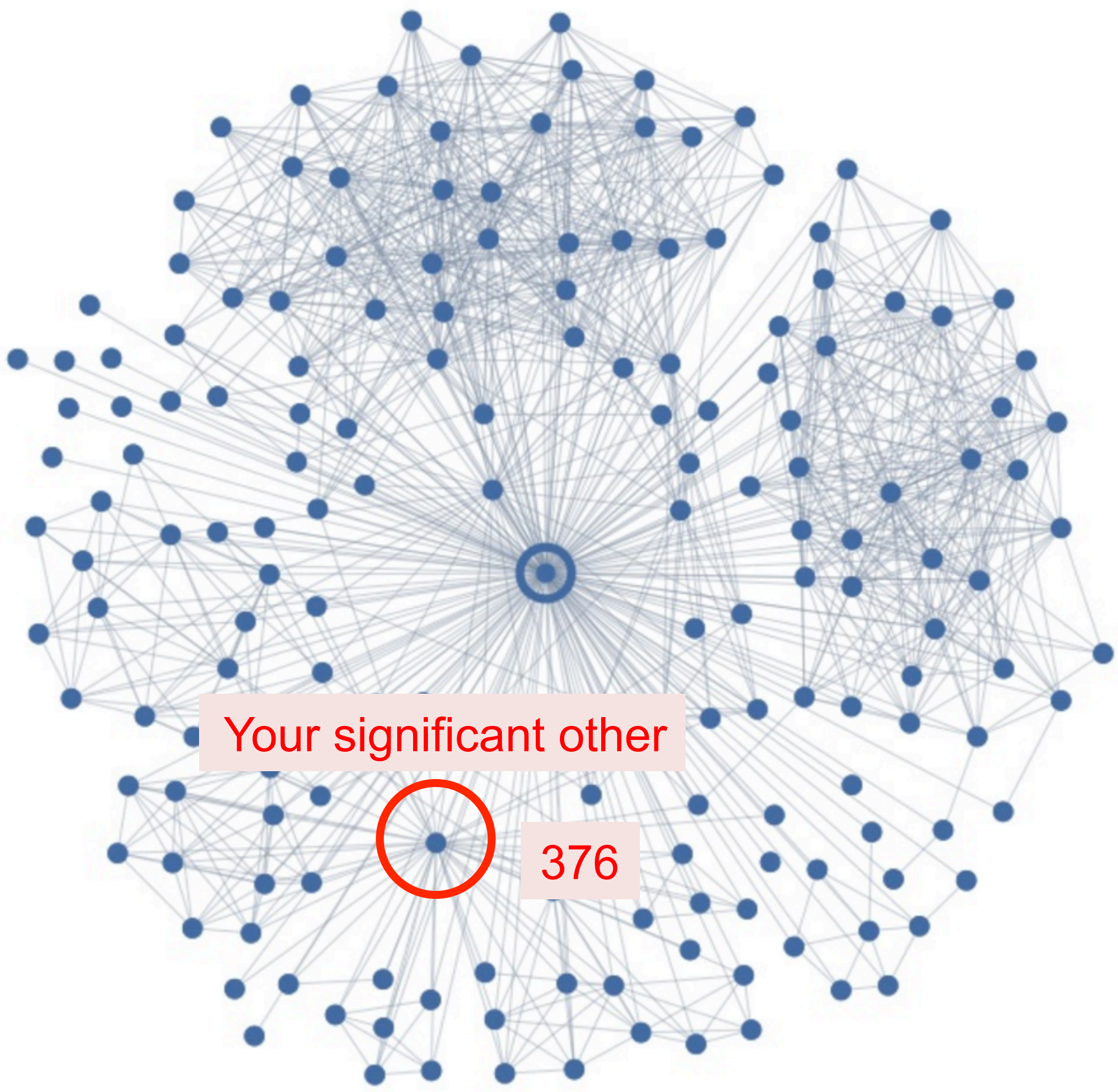
Who Do You Love?



Who Do You Love?



Who Do You Love?



Romance and Dispersion

Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook

Lars Backstrom
Facebook Inc.

Jon Kleinberg
Cornell University

ABSTRACT

A crucial task in the analysis of on-line social-networking systems is to identify important people — those linked by strong social ties — within an individual's network neighborhood. Here we investigate this question for a particular category of strong ties, those involving spouses or romantic partners. We organize our analysis around a basic question: given all the connections among a person's friends, can you recognize his or her romantic partner from the network structure alone? Using data from a large sample of Facebook users, we find that this task can be accomplished with high accuracy, but doing so requires the development of a new measure of tie strength that we term 'dispersion' — the extent to which two people's mutual friends are not themselves well-connected. The results offer methods for identifying types of structurally significant people in on-line applications, and suggest a potential expansion of existing theories of tie strength.

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database applications—*Data mining*

Keywords: Social Networks; Romantic Relationships.

they see from friends [1], and organizing their neighborhood into conceptually coherent groups [23, 25].

Tie Strength.

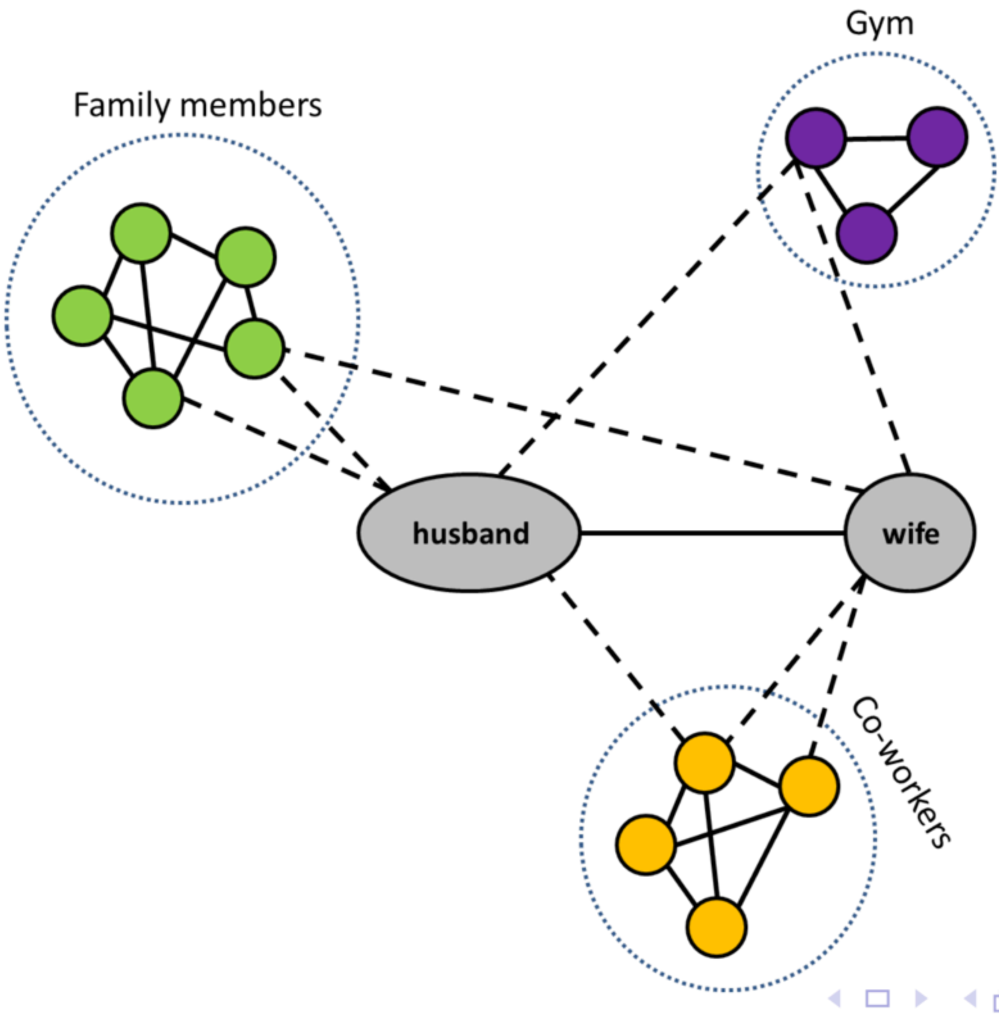
Tie strength forms an important dimension along which to characterize a person's links to their network neighbors. Tie strength informally refers to the 'closeness' of a friendship; it captures a spectrum that ranges from strong ties with close friends to weak ties with more distant acquaintances. An active line of research reaching back to foundational work in sociology has studied the relationship between the strengths of ties and their structural role in the underlying social network [15]. Strong ties are typically 'embedded' in the network, surrounded by a large number of mutual friends [6, 16], and often involving large amounts of shared time together [22] and extensive interaction [17]. Weak ties, in contrast, often involve few mutual friends and can serve as 'bridges' to diverse parts of the network, providing access to novel information [5, 15].

A fundamental question connected to our und
strong ties is to identify the mo

October 2013

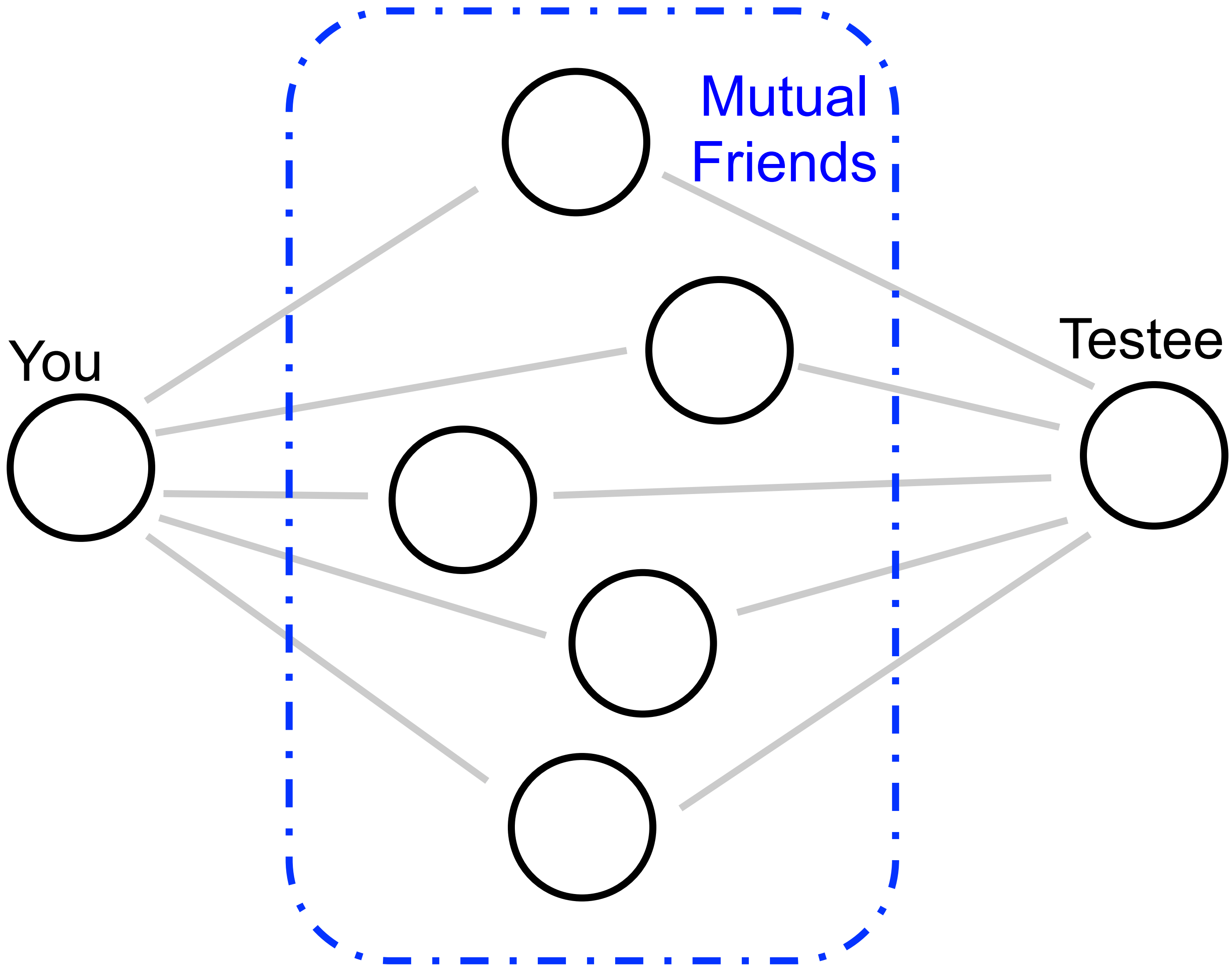
http://arxiv.org/pdf/1310.6753v1.pdf

Dispersion Insight



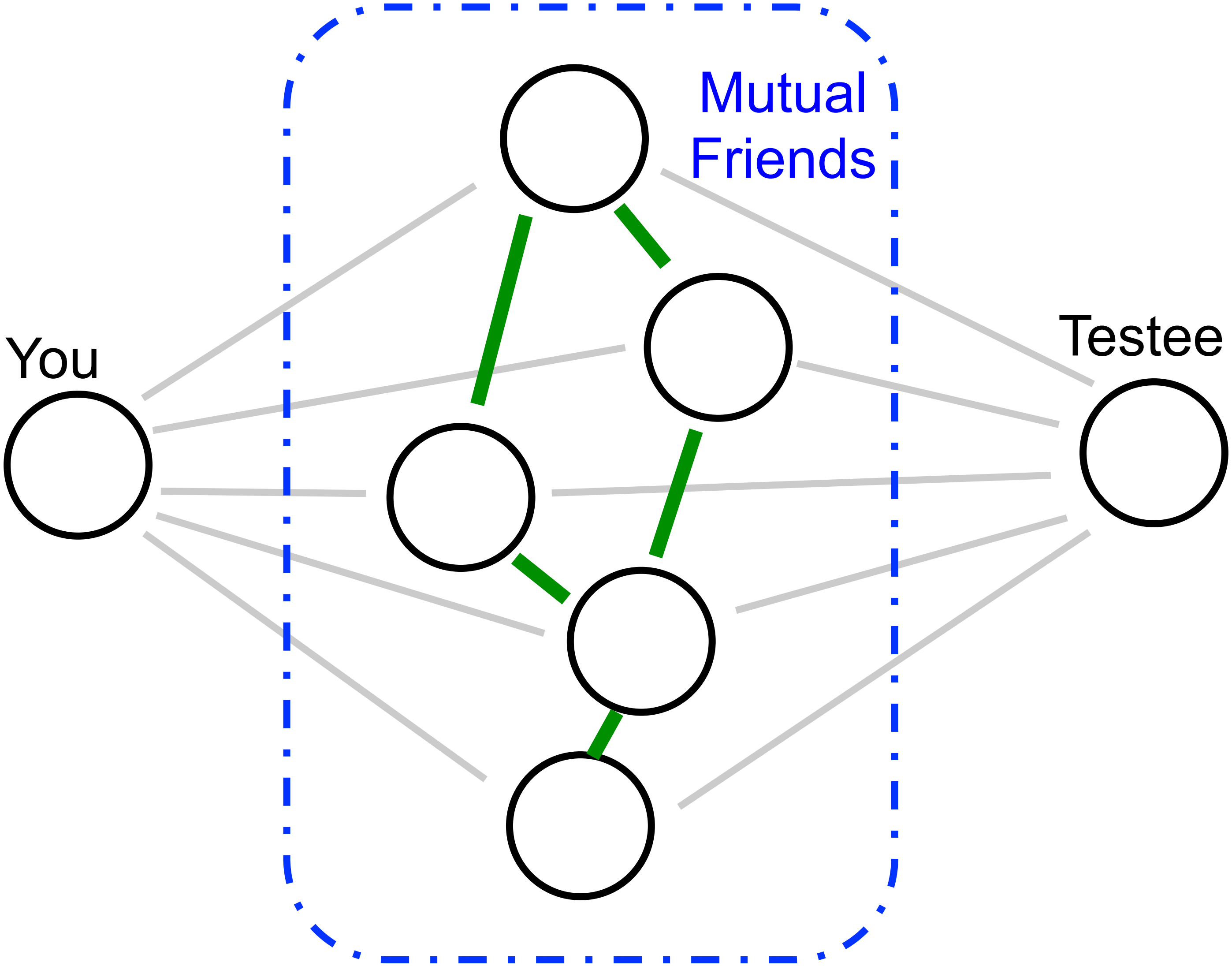
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



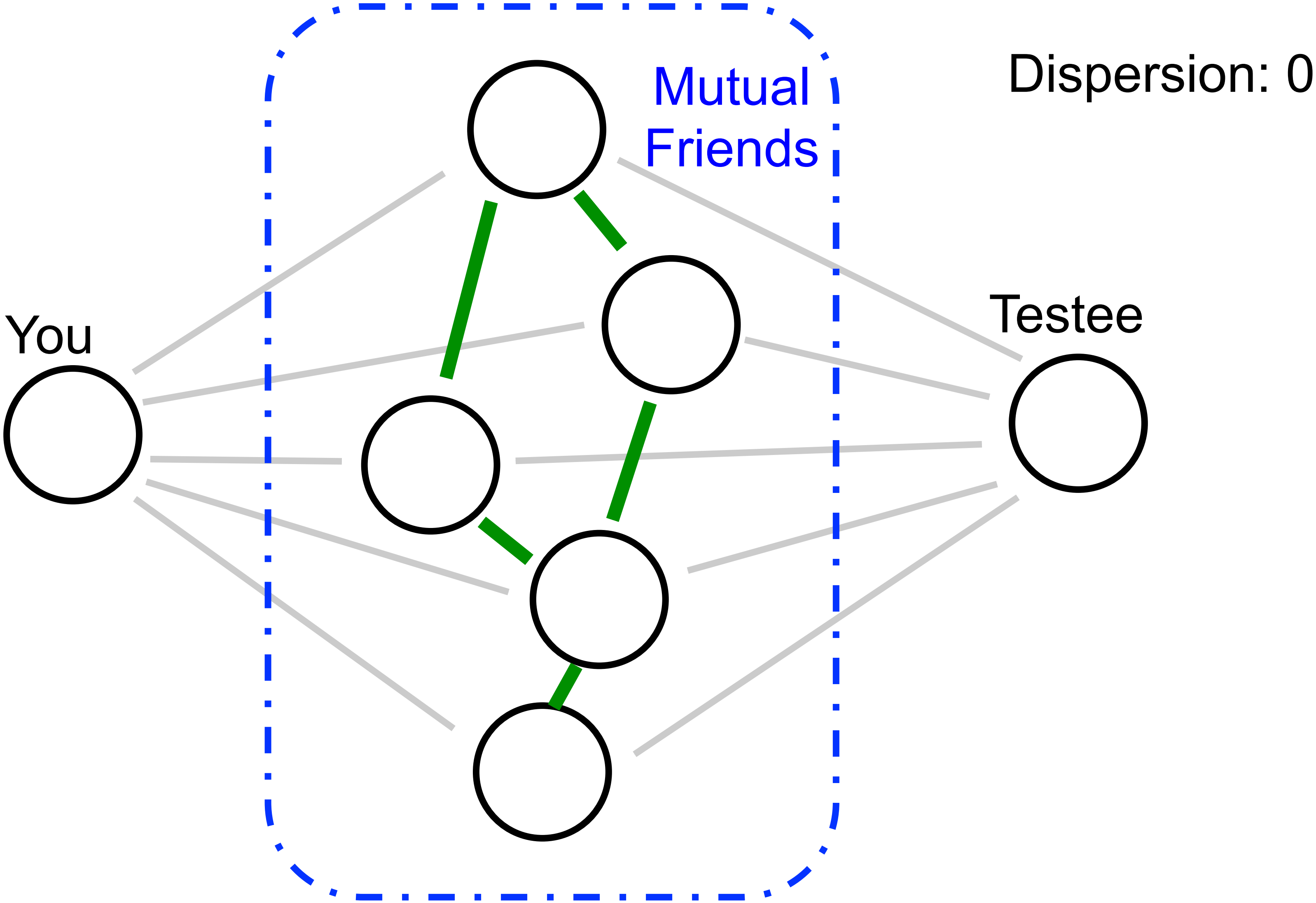
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



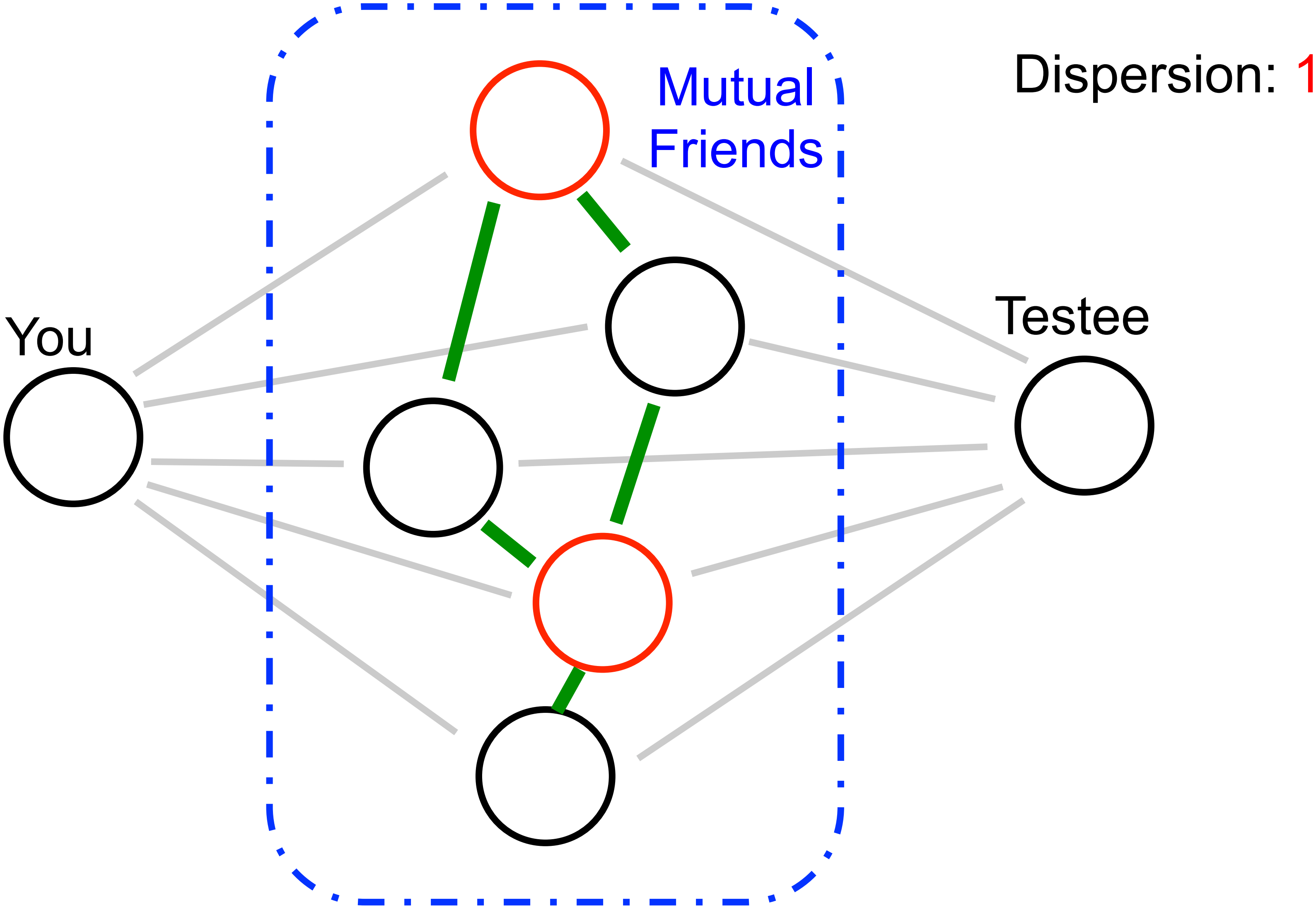
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



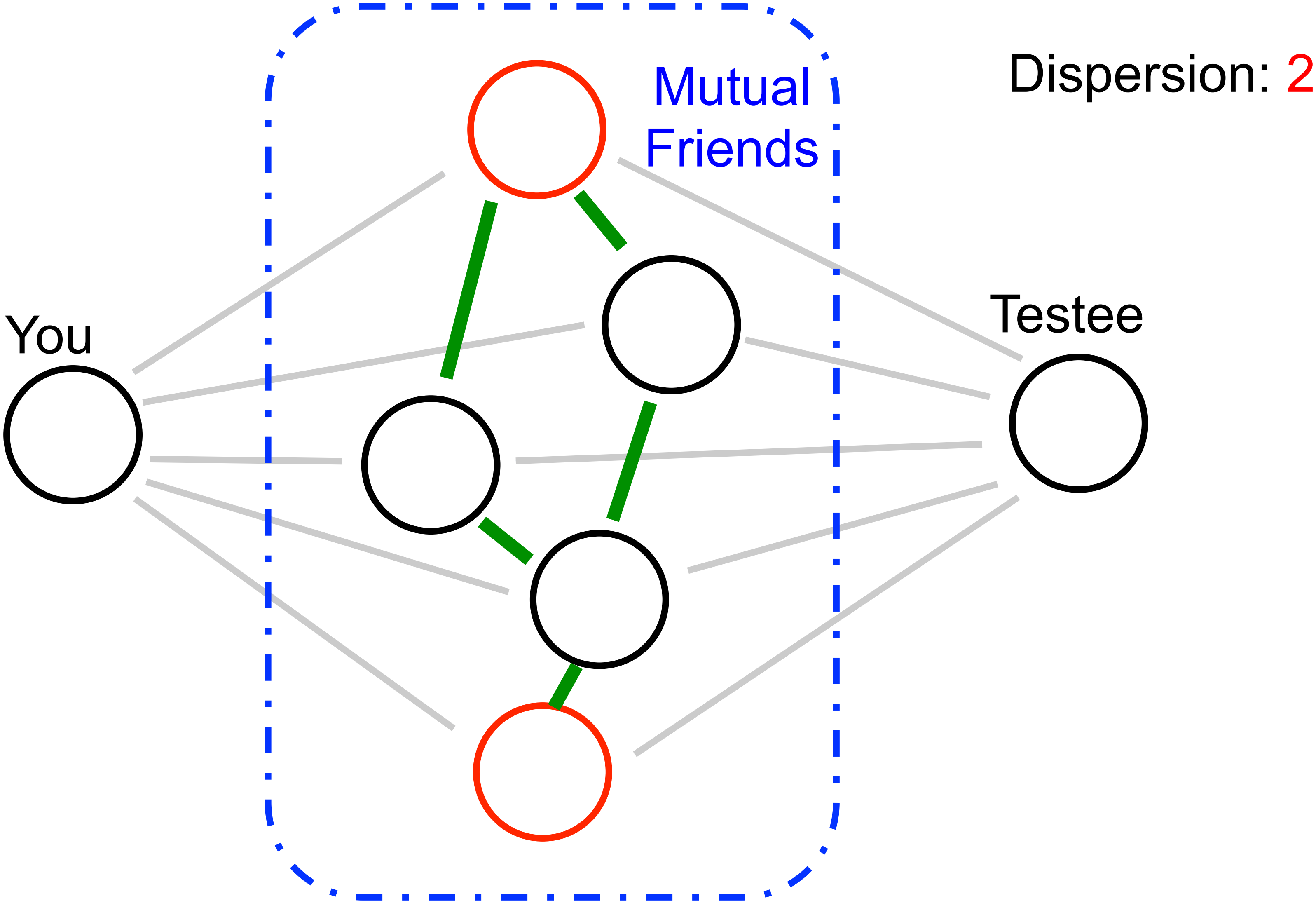
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



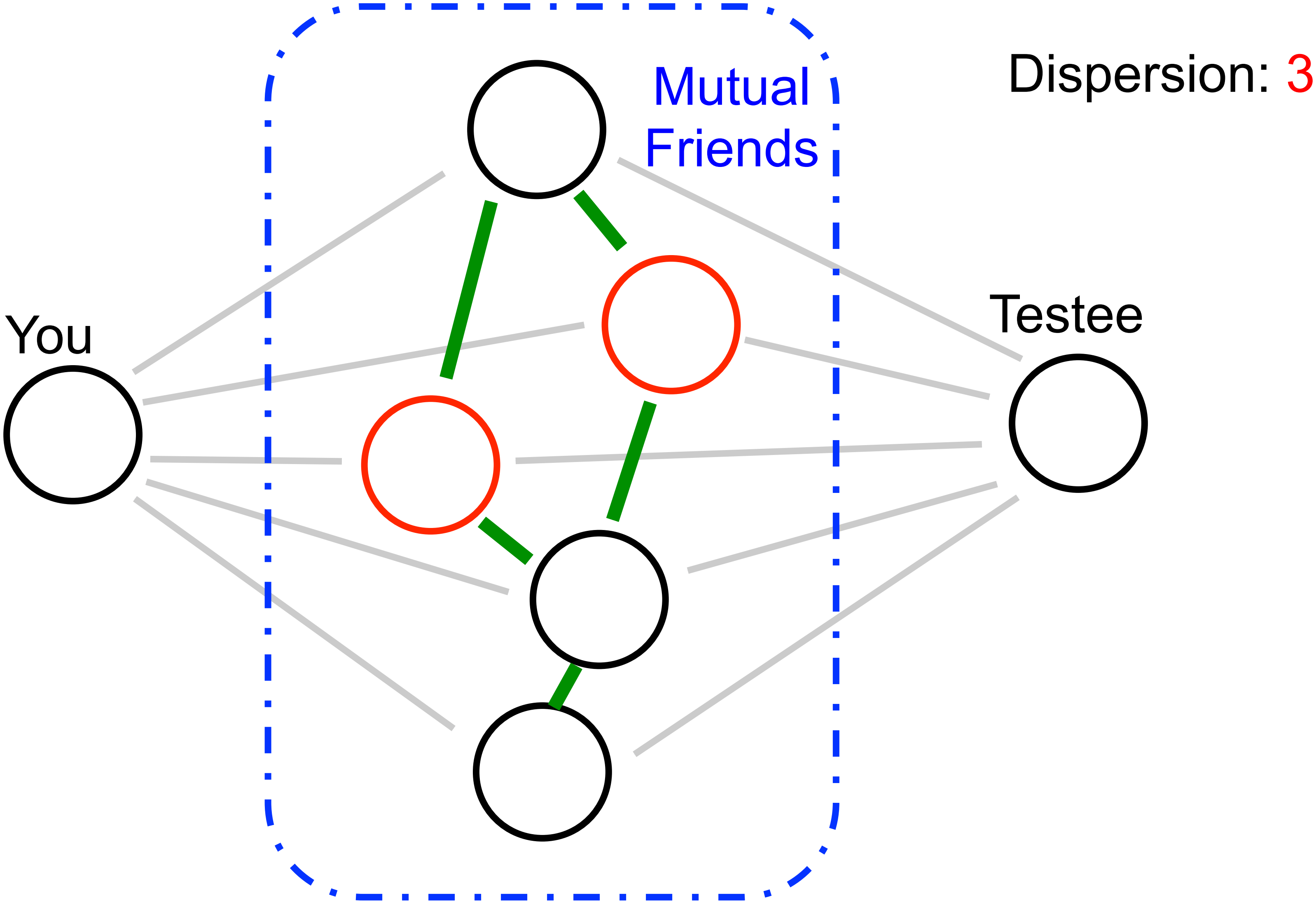
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



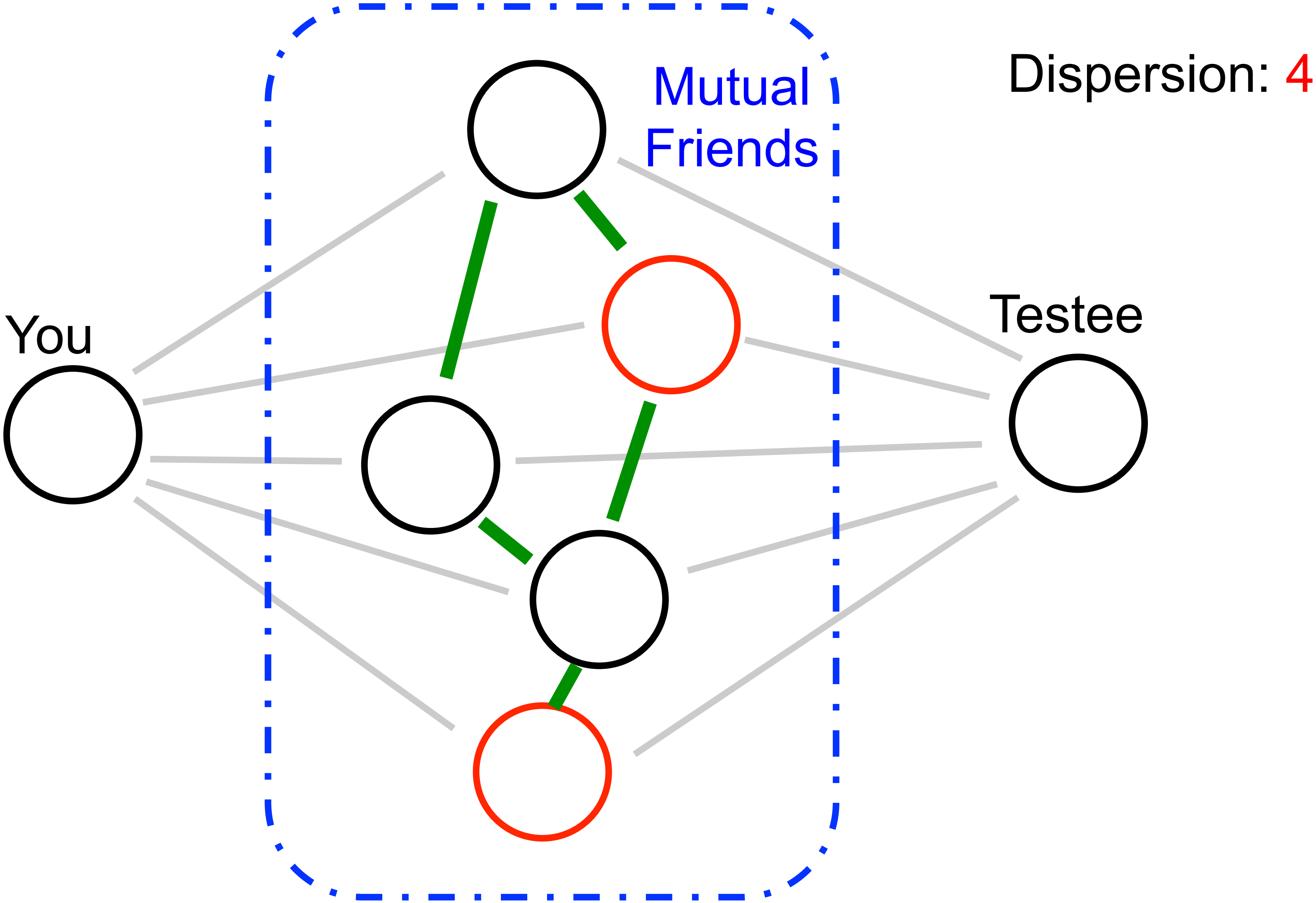
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



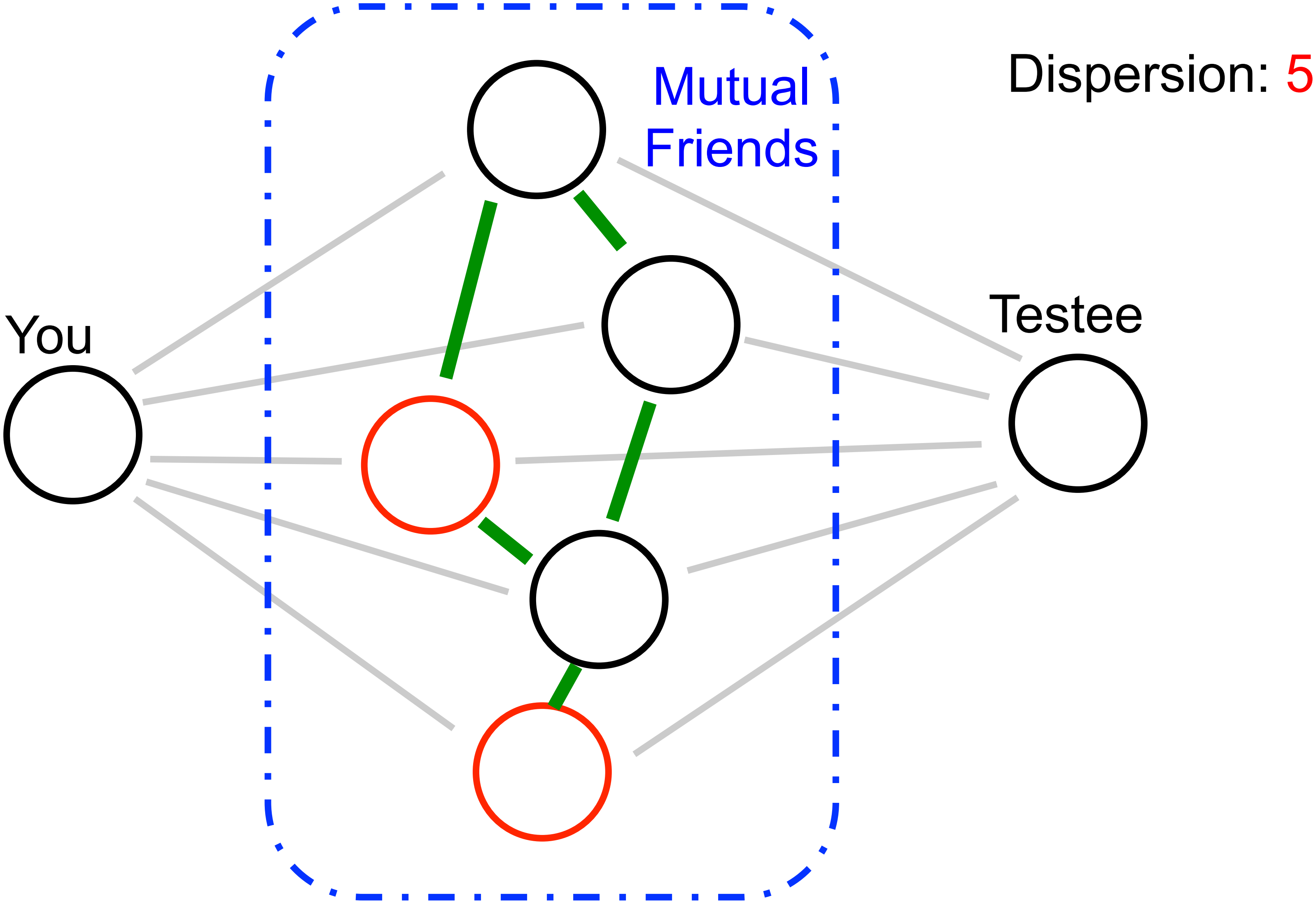
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



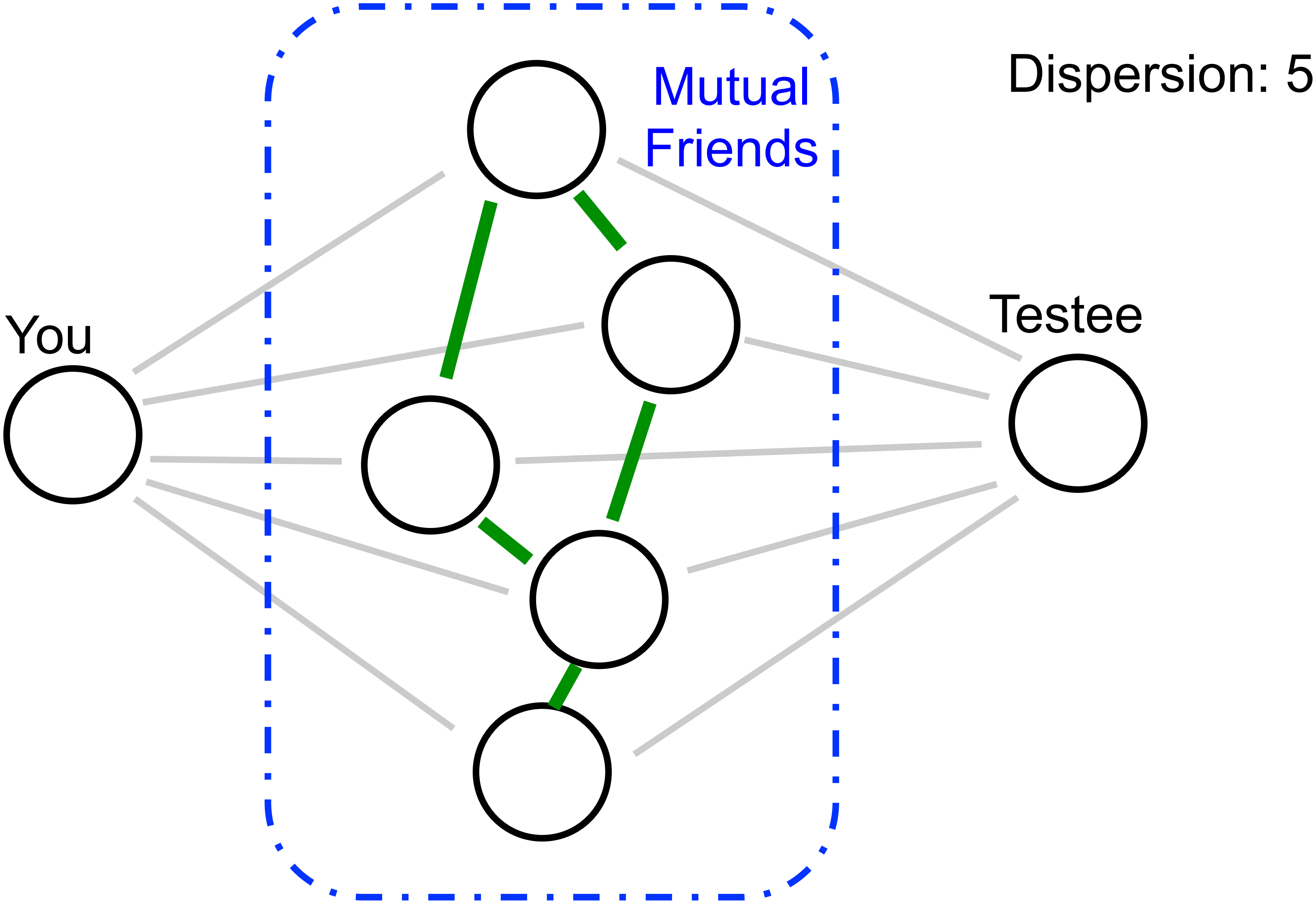
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion



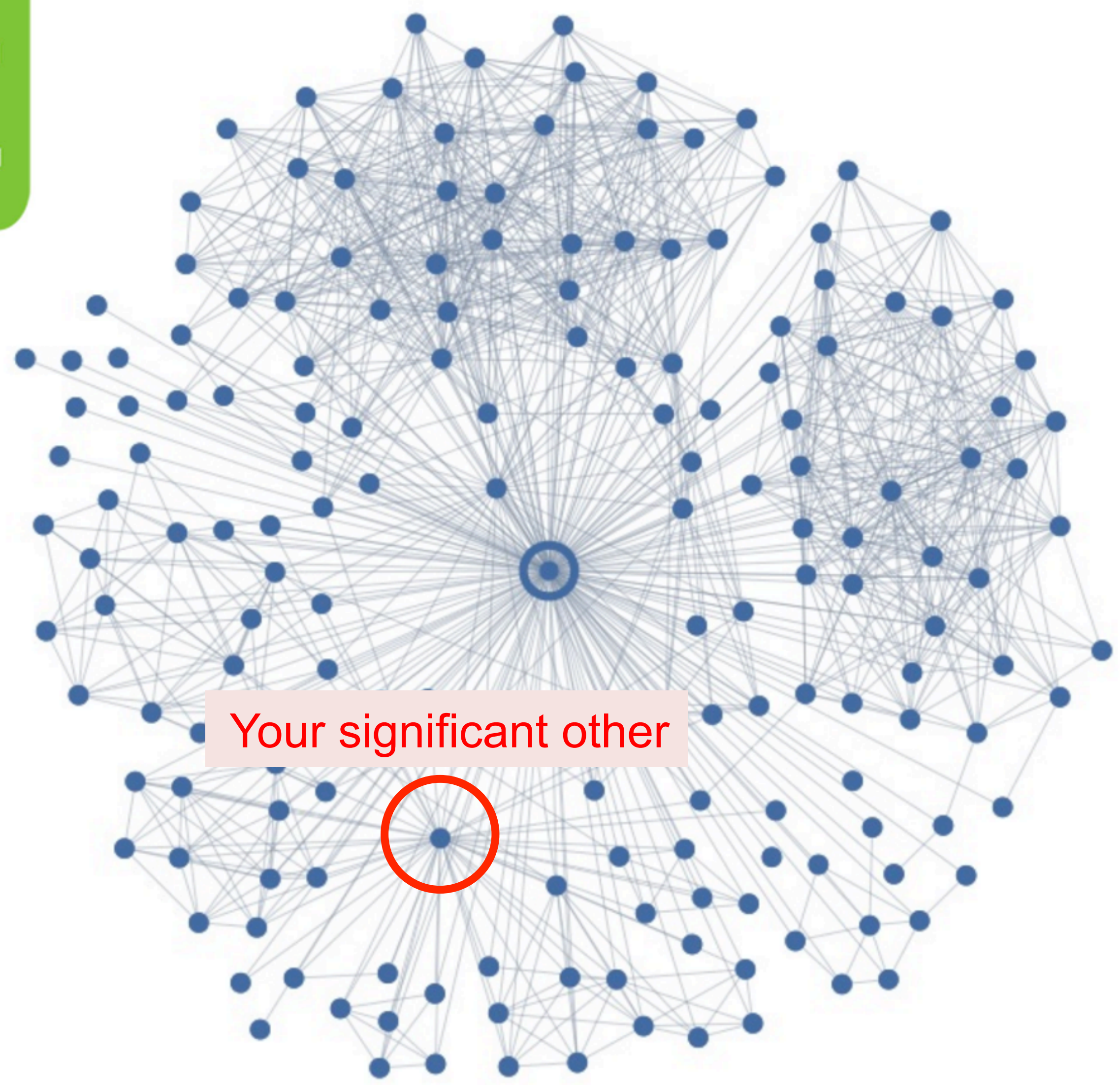
Dispersion: The extent to which two people's mutual friends are not directly connected

Dispersion

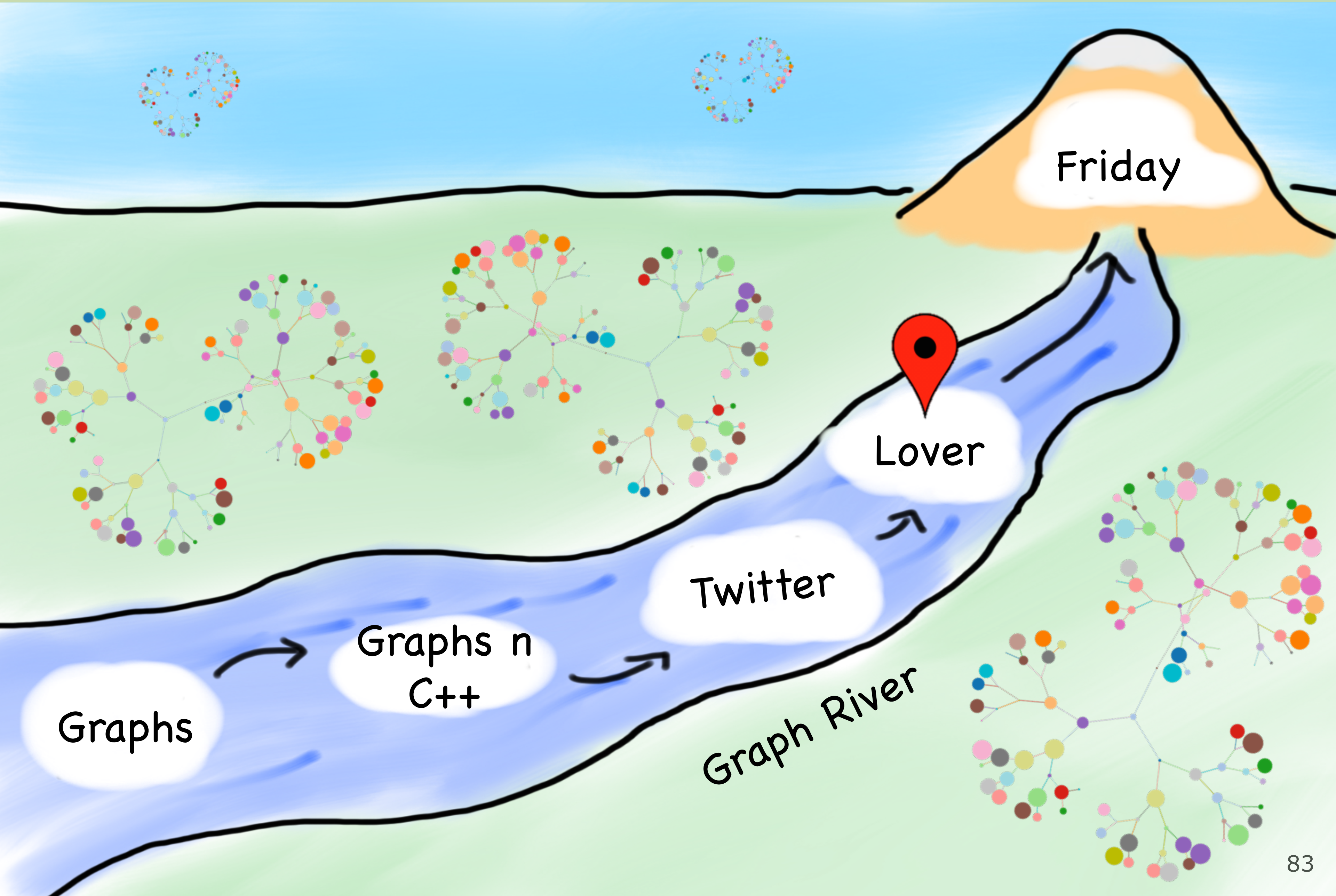


Dispersion: The extent to which two people's mutual friends are not directly connected

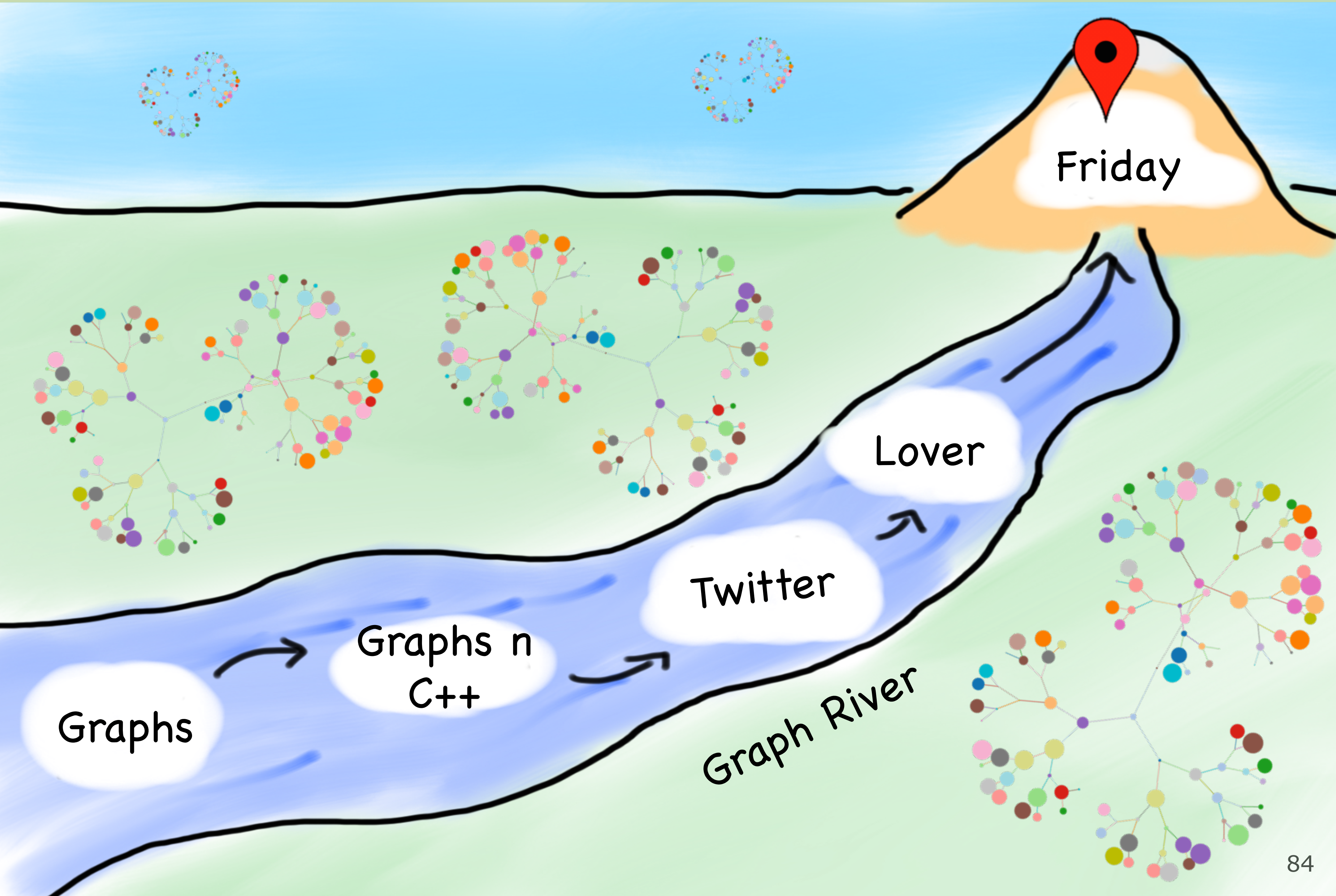
Who Do You Love?



Today's Route



Today's Route



Today's Goal

1. Introduction to Graphs
2. Graphs in C++

Page Rank