Graphs and Networks

Announcements

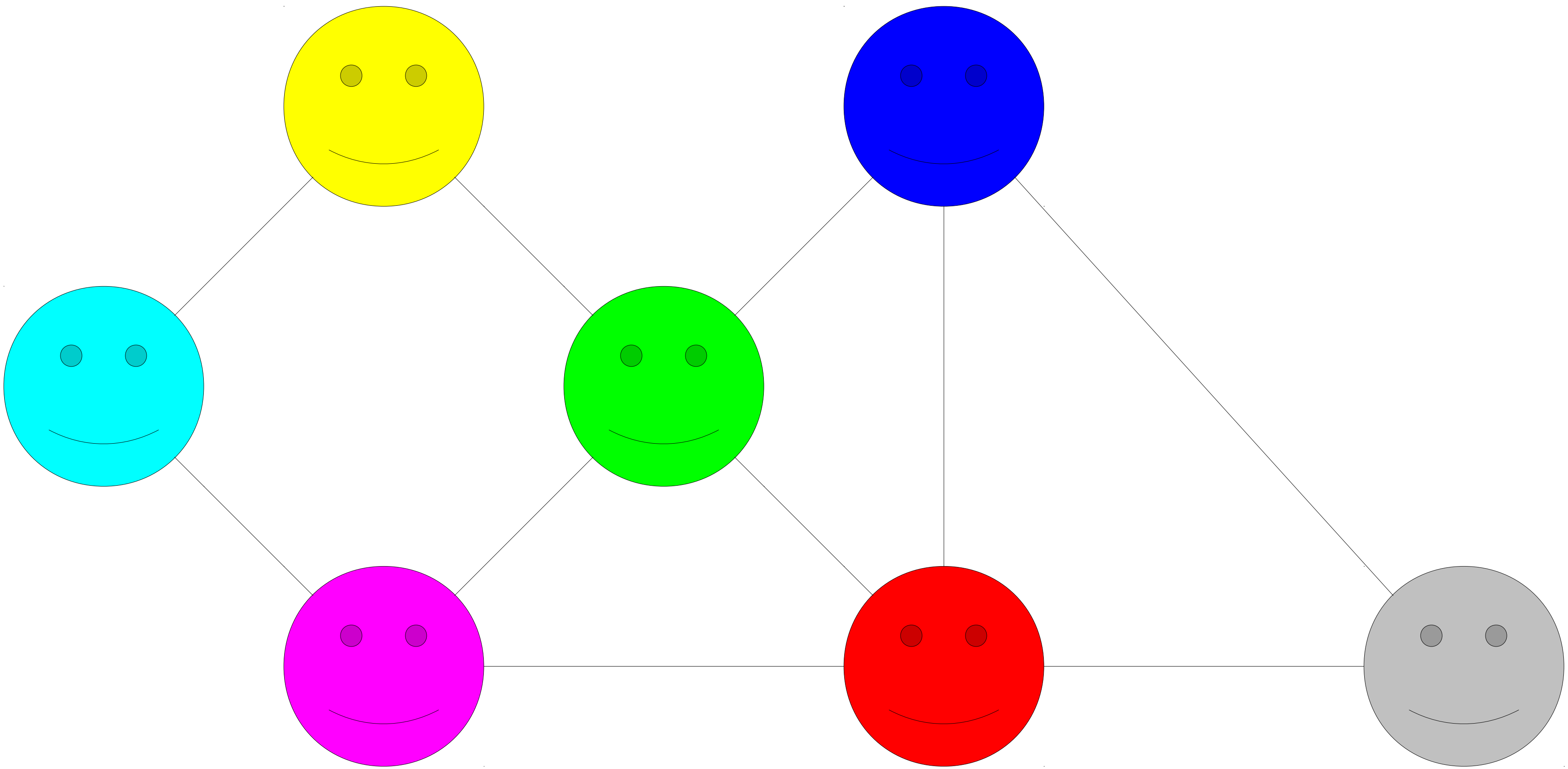
- Casual CS Dinner for Women Studying Computer Science: **Thursday, March 7** at **6PM** in **Gates 219**!
- RSVP through the email link sent out earlier today.

Announcements

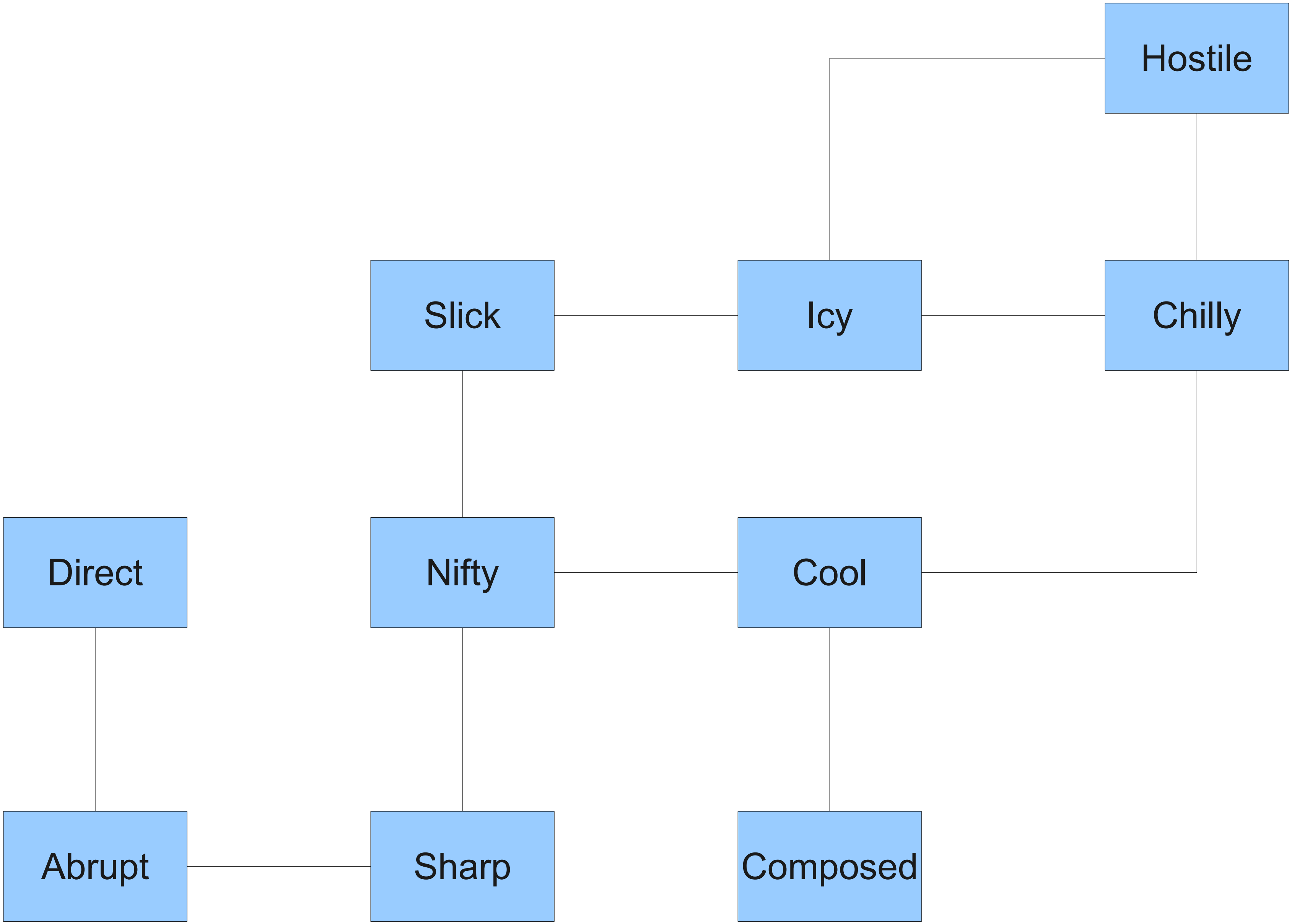
- Assignment 5 due right now.
- Assignment 6 (**NameSurfer**) out, due next **Wednesday, March 13** at 3:15PM
- Second Midterm exam next Monday, March 11 from 7PM – 10PM in MemAud.
  - Covers material up through and including Wednesday's lecture.
  - Practice exam released today; solutions go out on Wednesday.
  - Email Gil **no later than 11:59PM** on Wednesday if you need to take the exam at an alternate time.

NameSurfer Demo

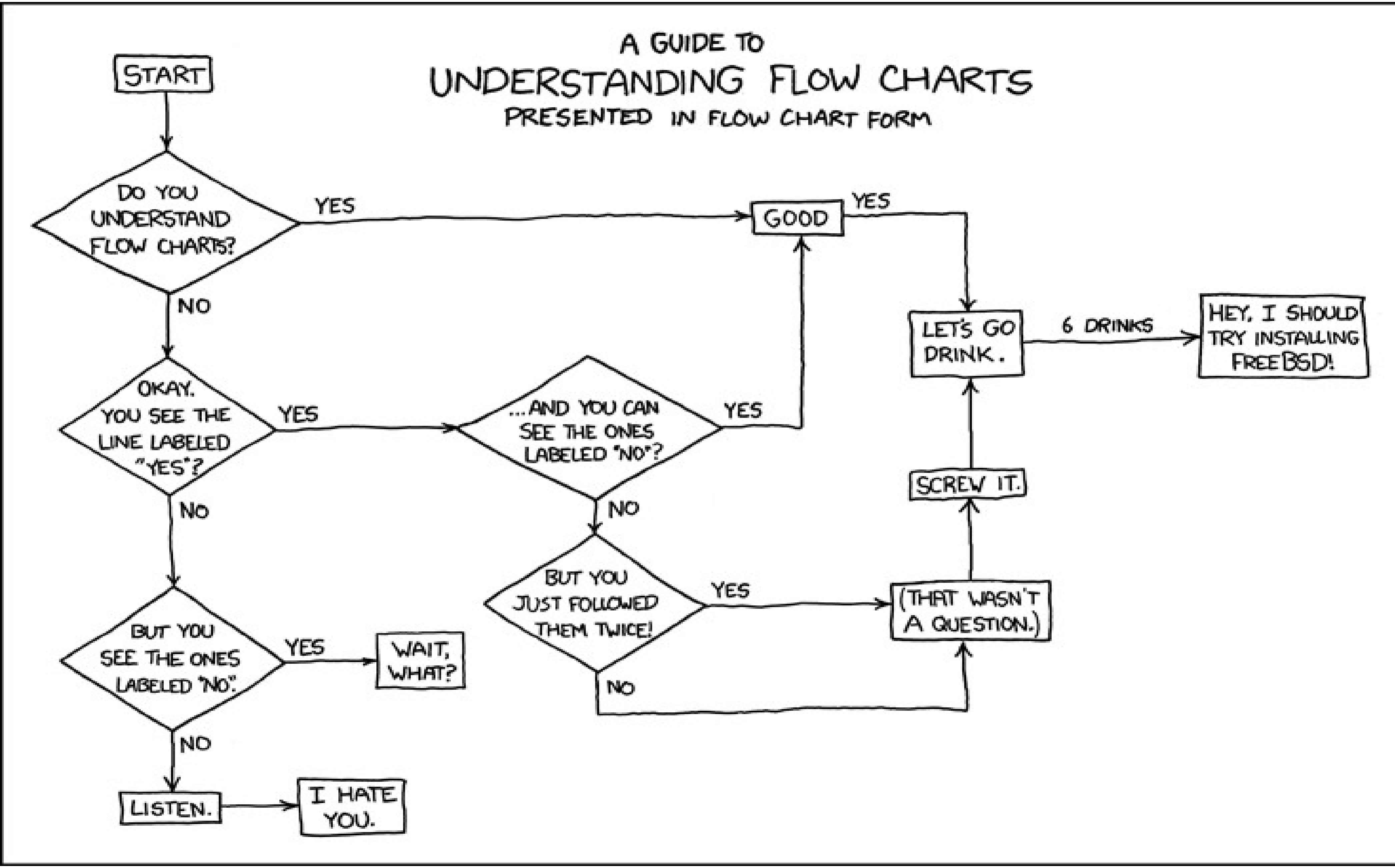
A Social Network



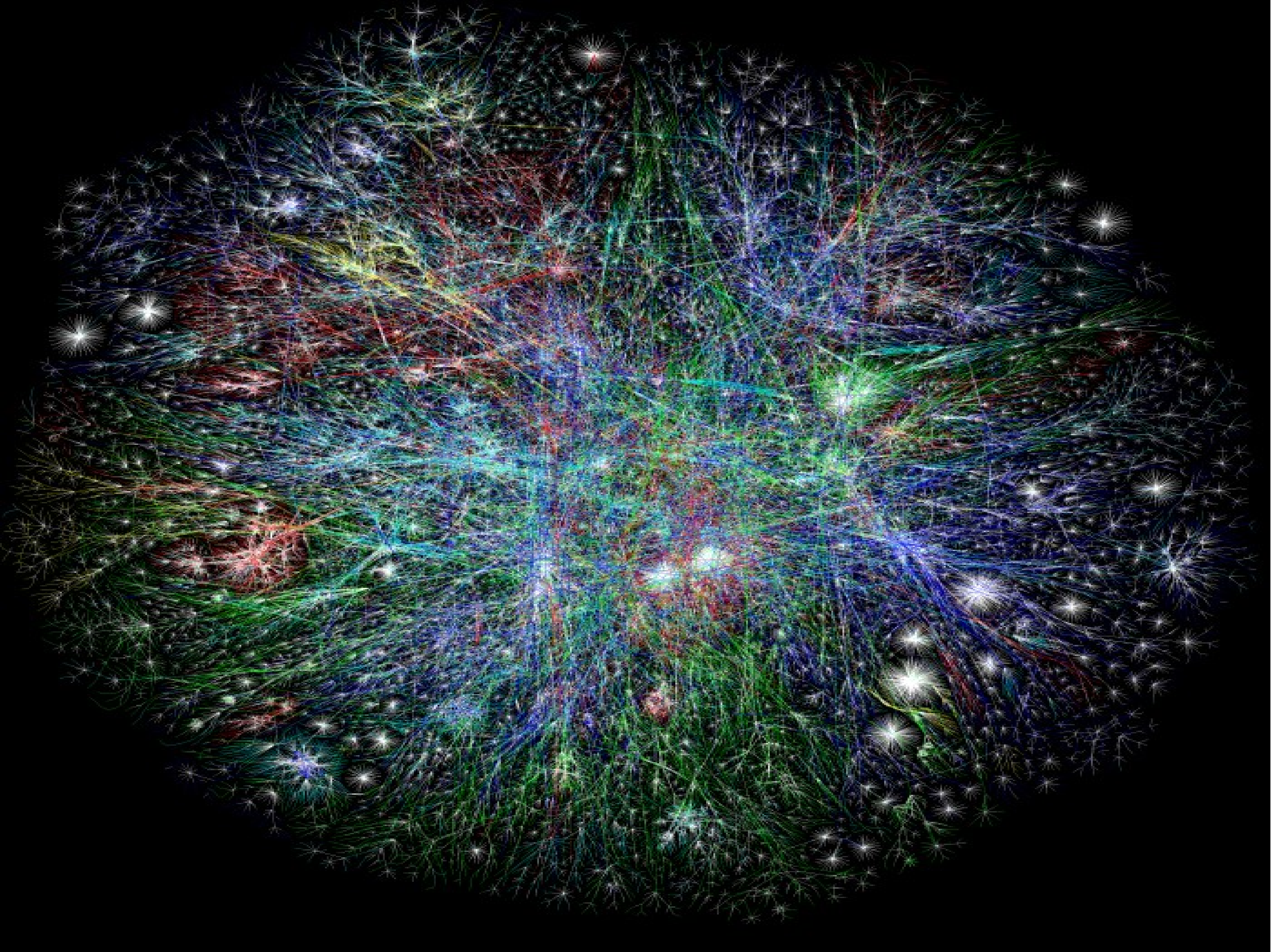
Synonyms



Hostile
Slick
Icy
Chilly
Direct
Nifty
Cool
Abrupt
Sharp
Composed


A GUIDE TO
UNDERSTANDING FLOW CHARTS
PRESENTED IN FLOW CHART FORM
START
DO YOU UNDERSTAND FLOW CHARTS?
YES
GOOD
YES
NO
LET'S GO DRINK.
6 DRINKS
HEY, I SHOULD TRY INSTALLING FREEBSD!
OKAY. YOU SEE THE LINE LABELED "YES"?
YES
...AND YOU CAN SEE THE ONES LABELED "NO"?
YES
NO
NO
SCREW IT.
BUT YOU JUST FOLLOWED THEM TWICE!
YES
(THAT WASN'T A QUESTION.)
NO
BUT YOU SEE THE ONES LABELED "NO".
YES
WAIT, WHAT?
NO
LISTEN.
I HATE YOU.

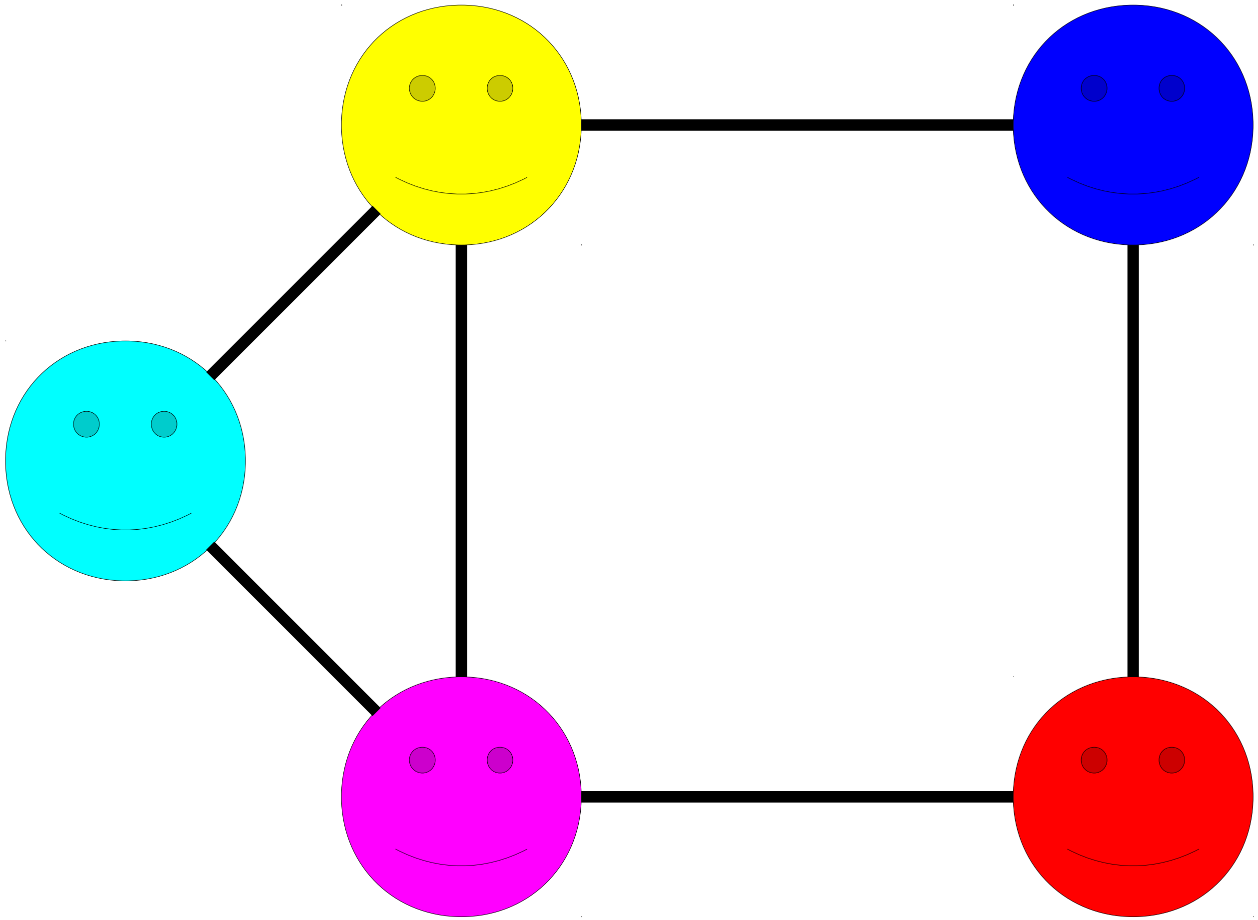
Source: xkcd



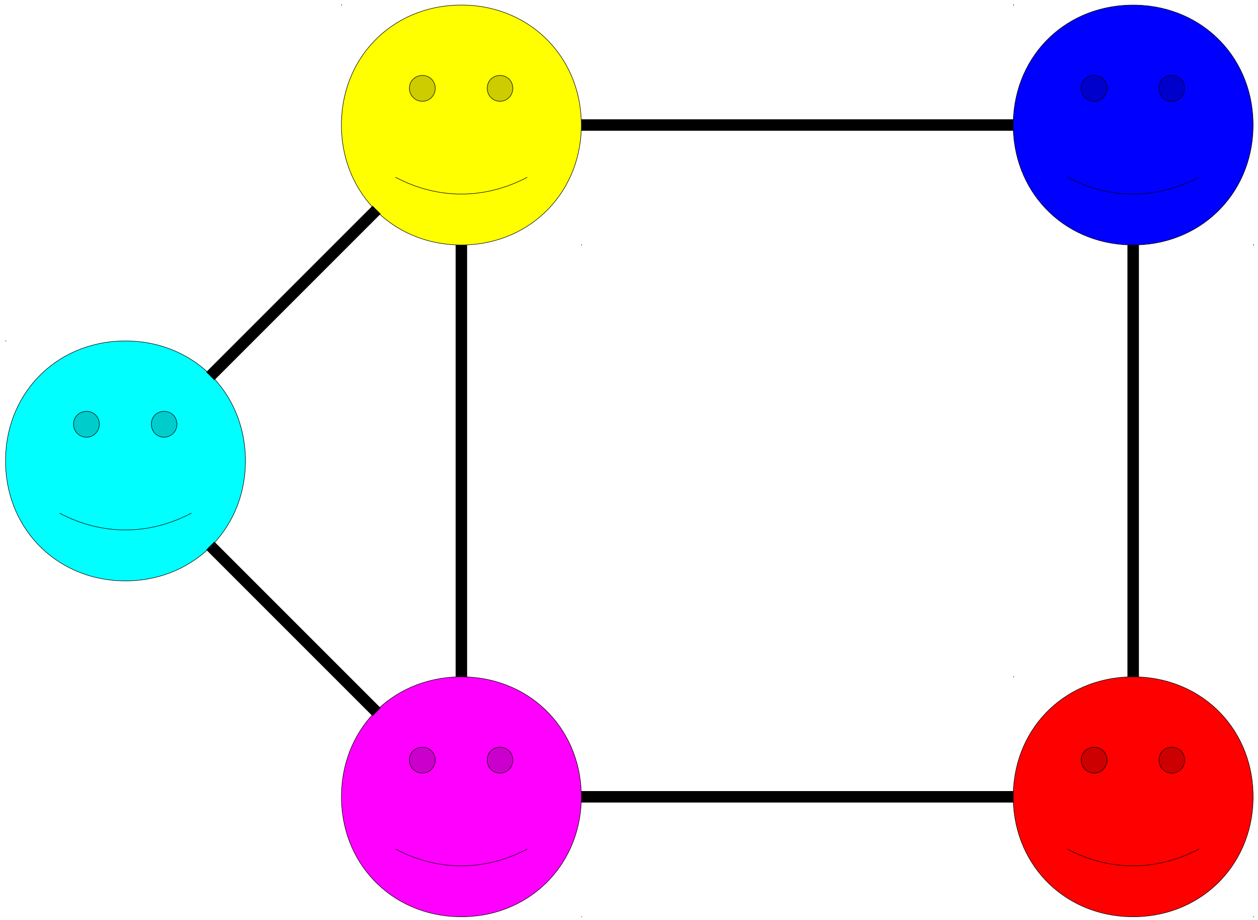
A **graph** is a mathematical structure for representing relationships.

A **graph** is a mathematical structure for representing relationships.

A **graph** is a mathematical structure for representing relationships.

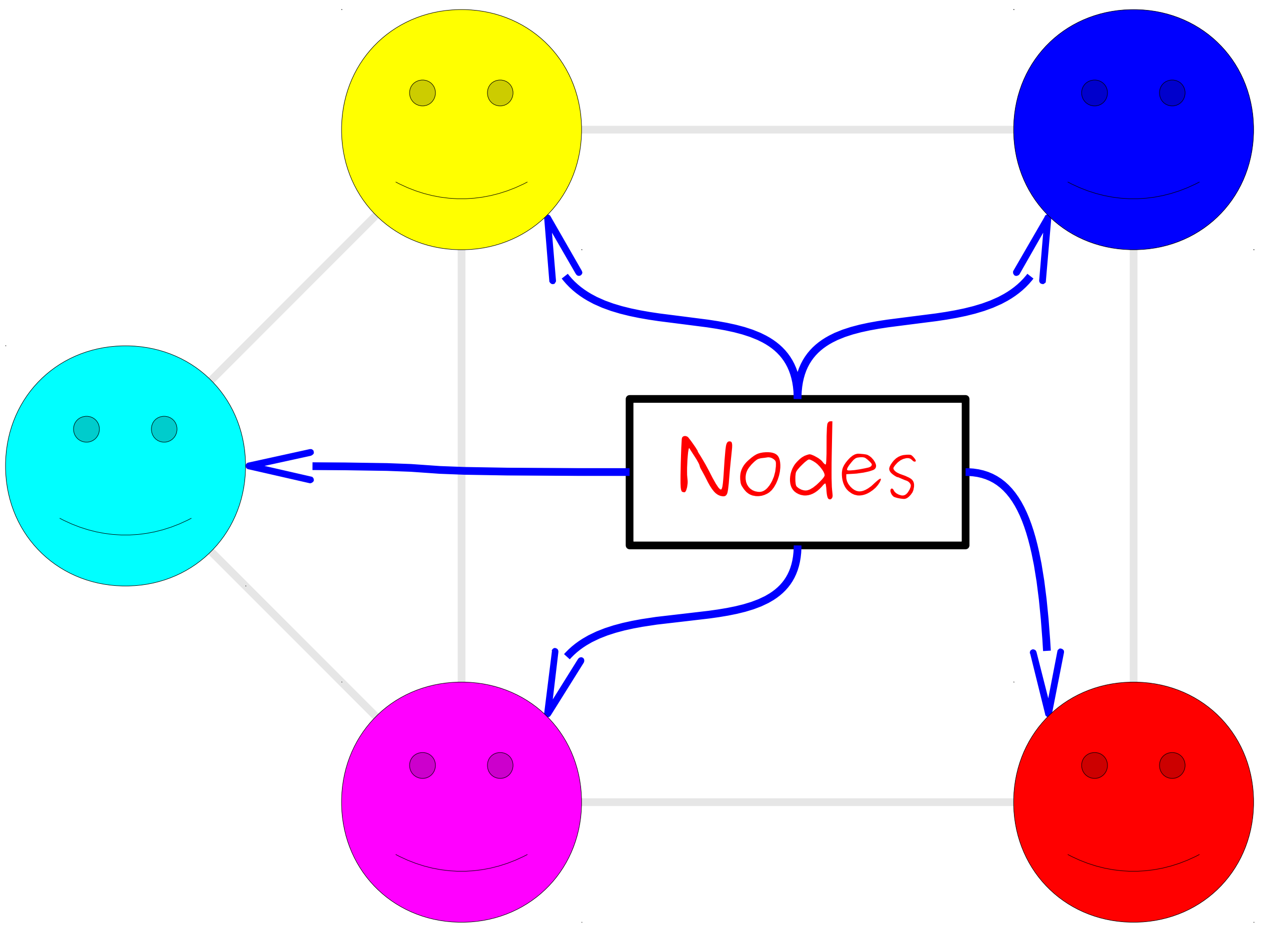


A **graph** is a mathematical structure for representing relationships.



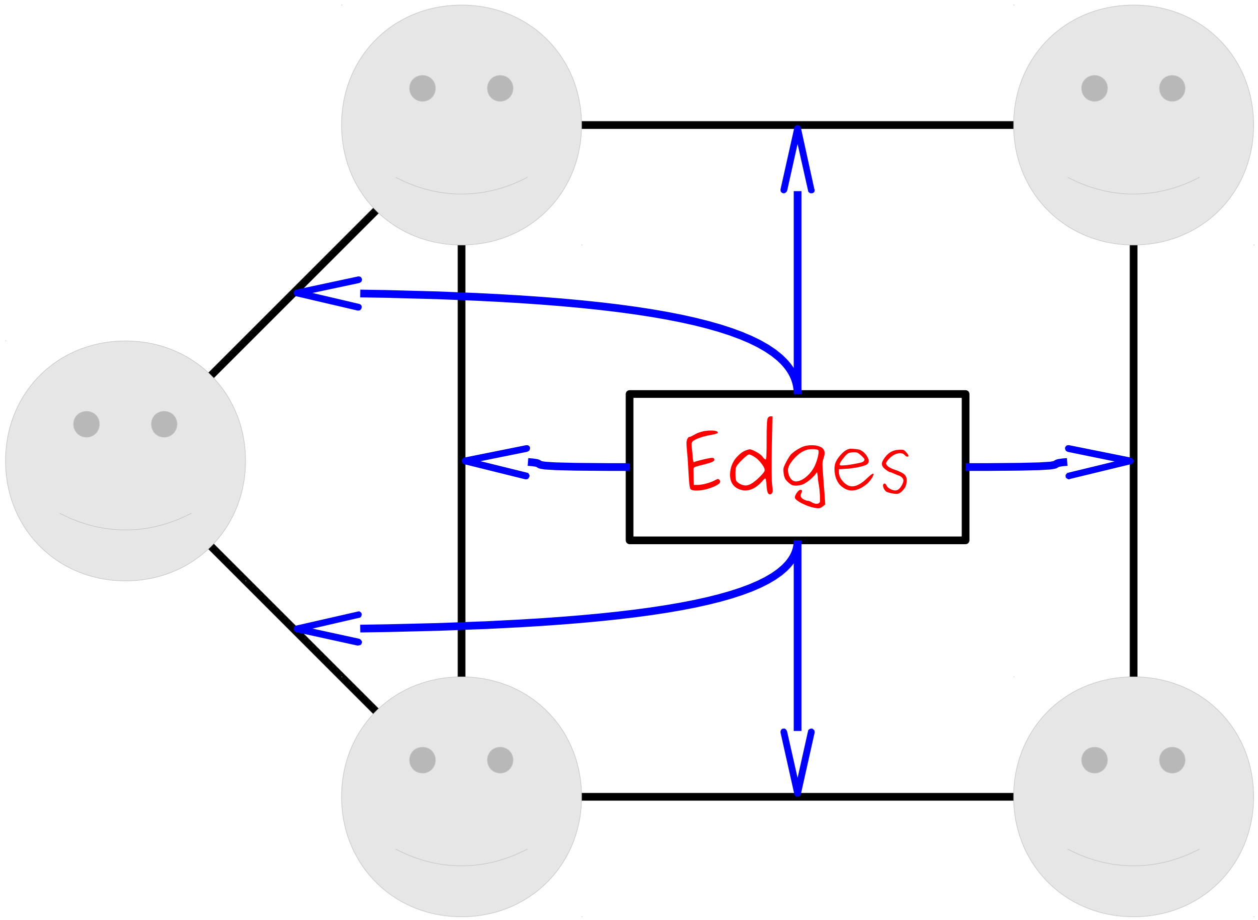
A graph consists of a set of **nodes** connected by **edges**.

A **graph** is a mathematical structure for representing relationships.



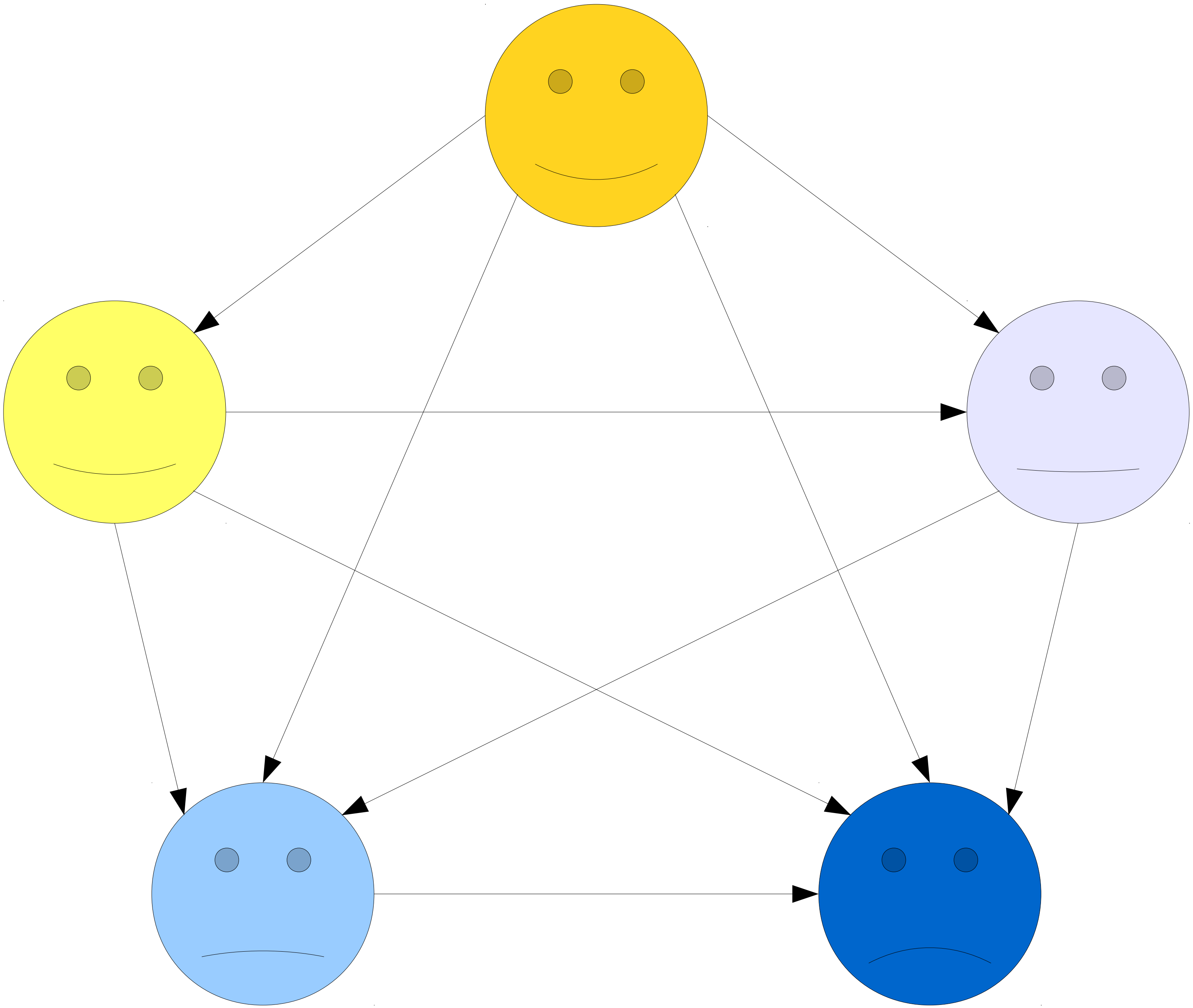
A graph consists of a set of **nodes** connected by **edges**.

A **graph** is a mathematical structure for representing relationships.

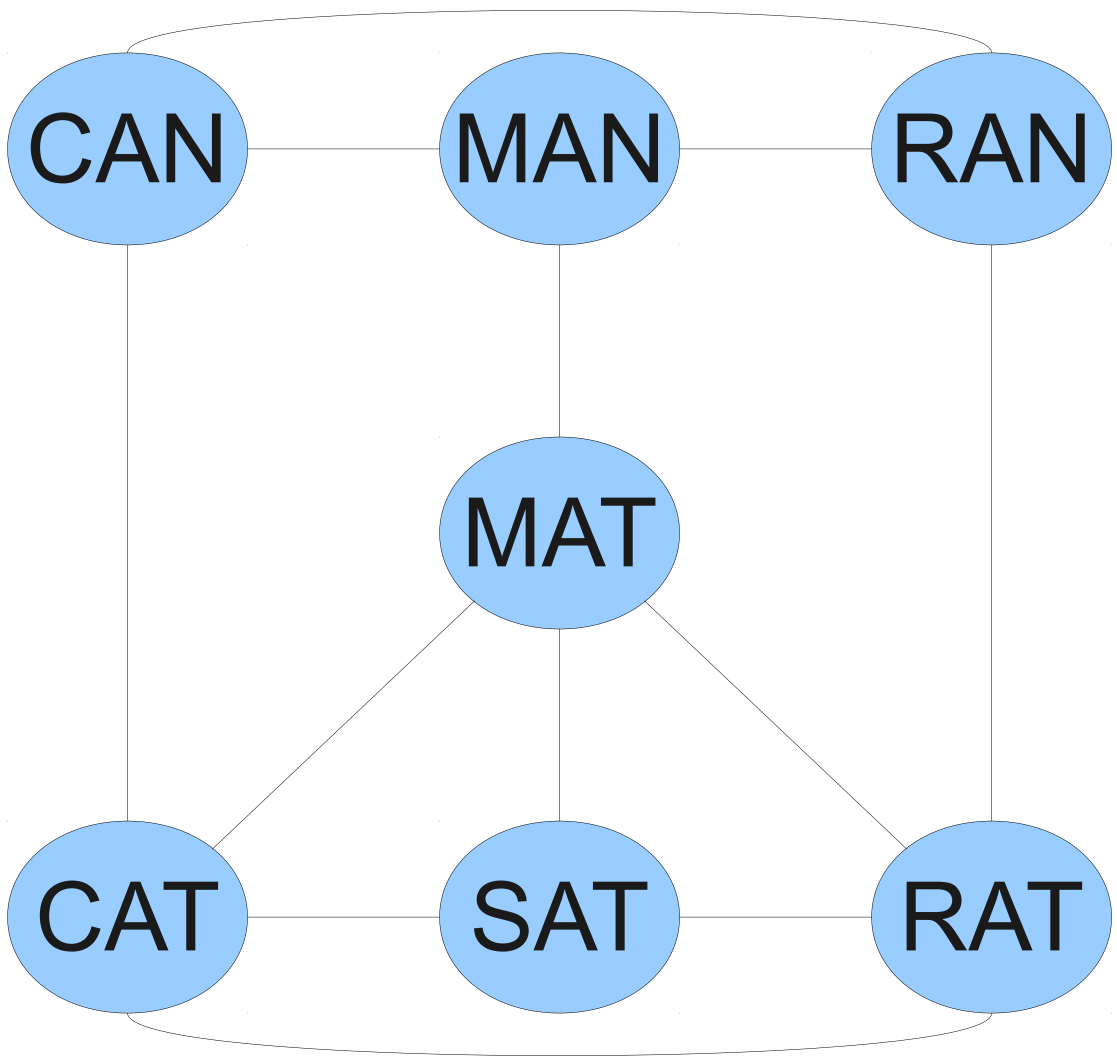


A graph consists of a set of **nodes** connected by **edges**.

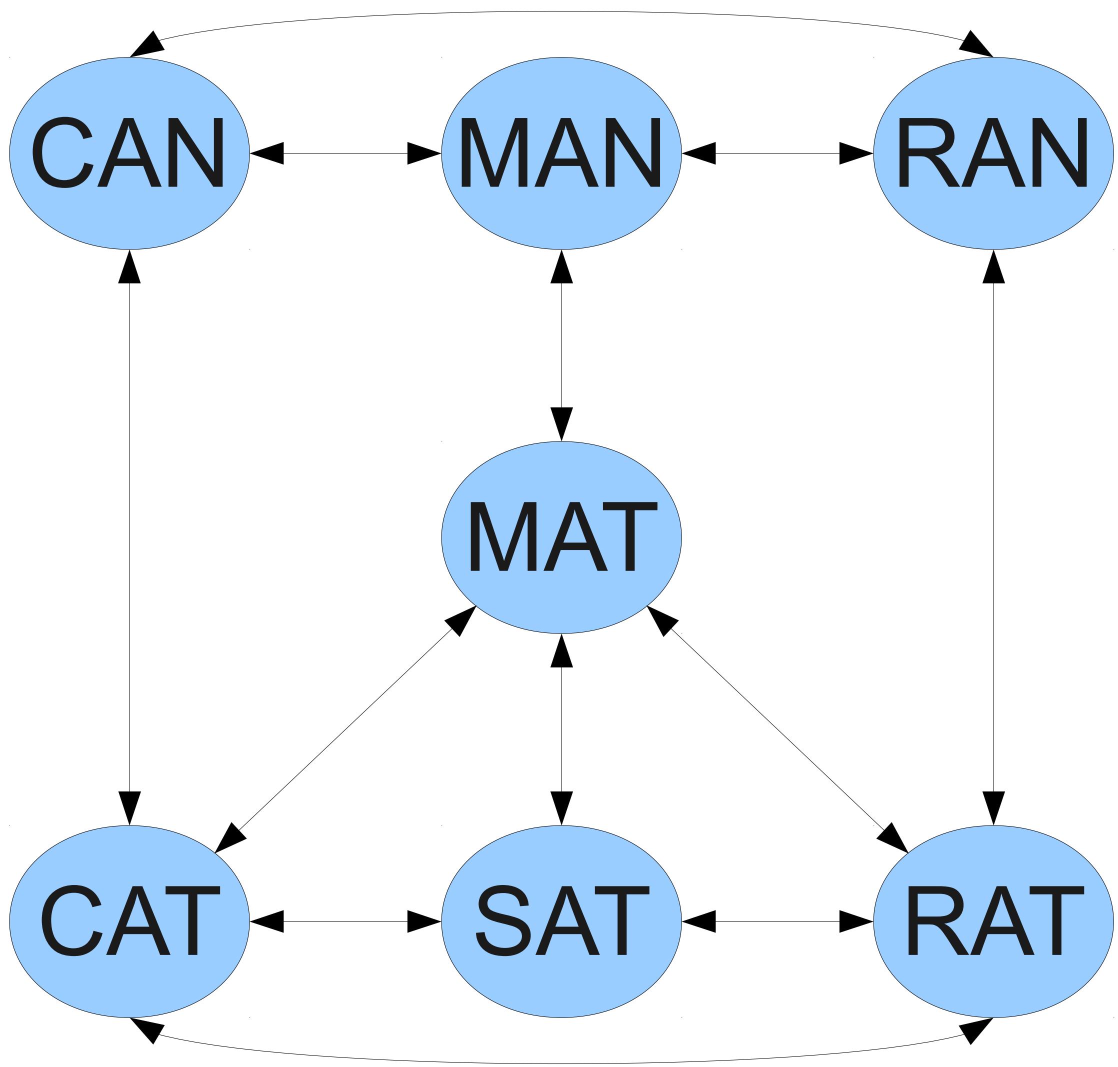
Some graphs are **directed**.



Some graphs are **undirected**.



Some graphs are **undirected**.



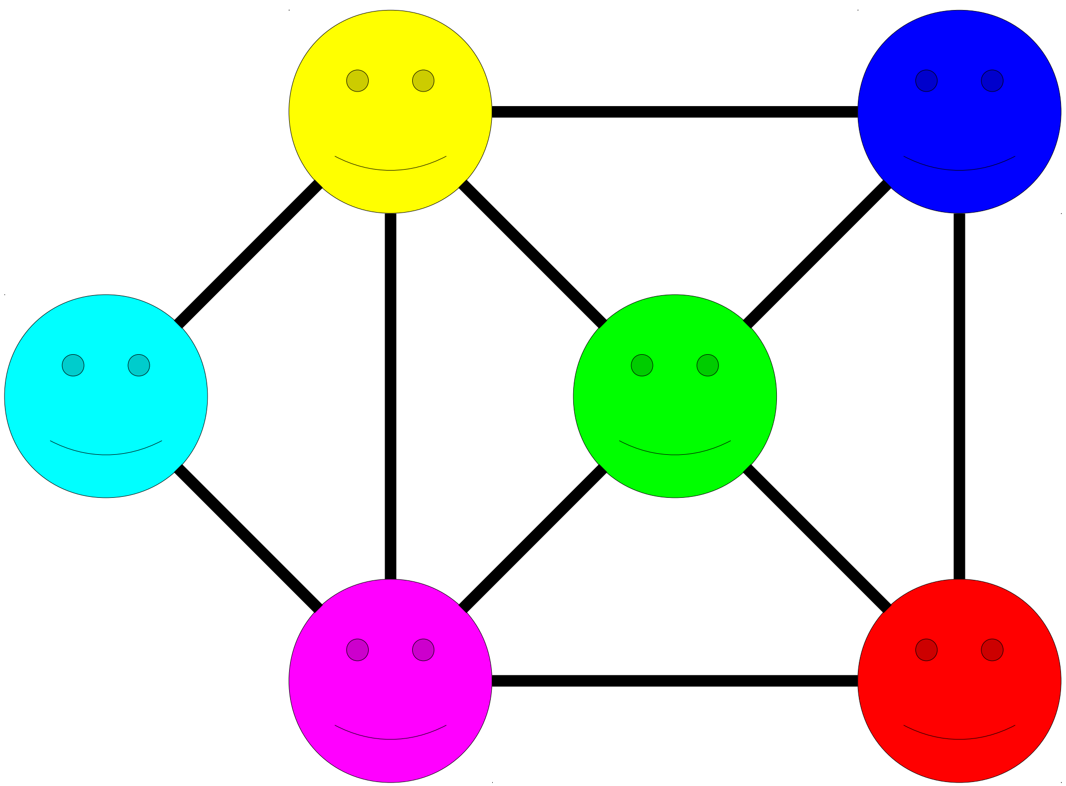
You can think of them as directed graphs with edges both ways.

How can we represent graphs in Java?

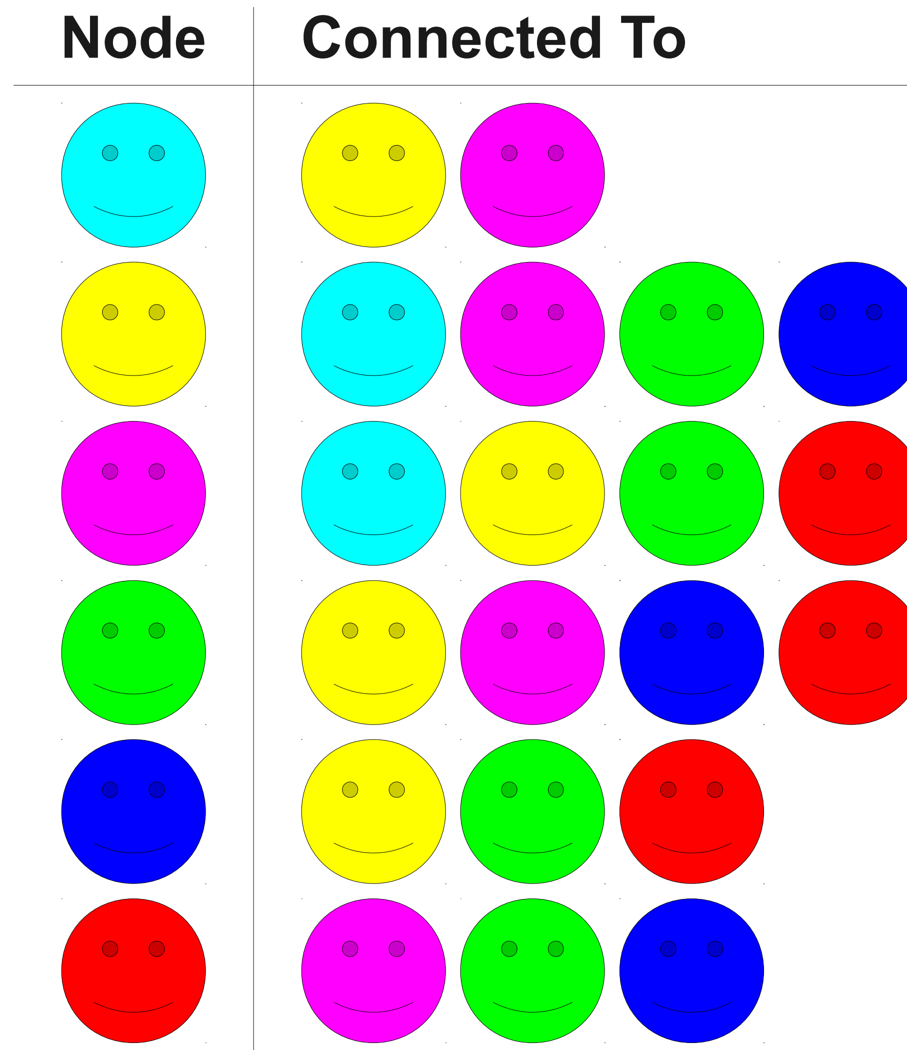
Representing Graphs

We can represent a graph as a map from nodes to the list of nodes each node is connected to.

`HashMap<Node, ArrayList<Node>>`



`Node` `ArrayList<Node>`



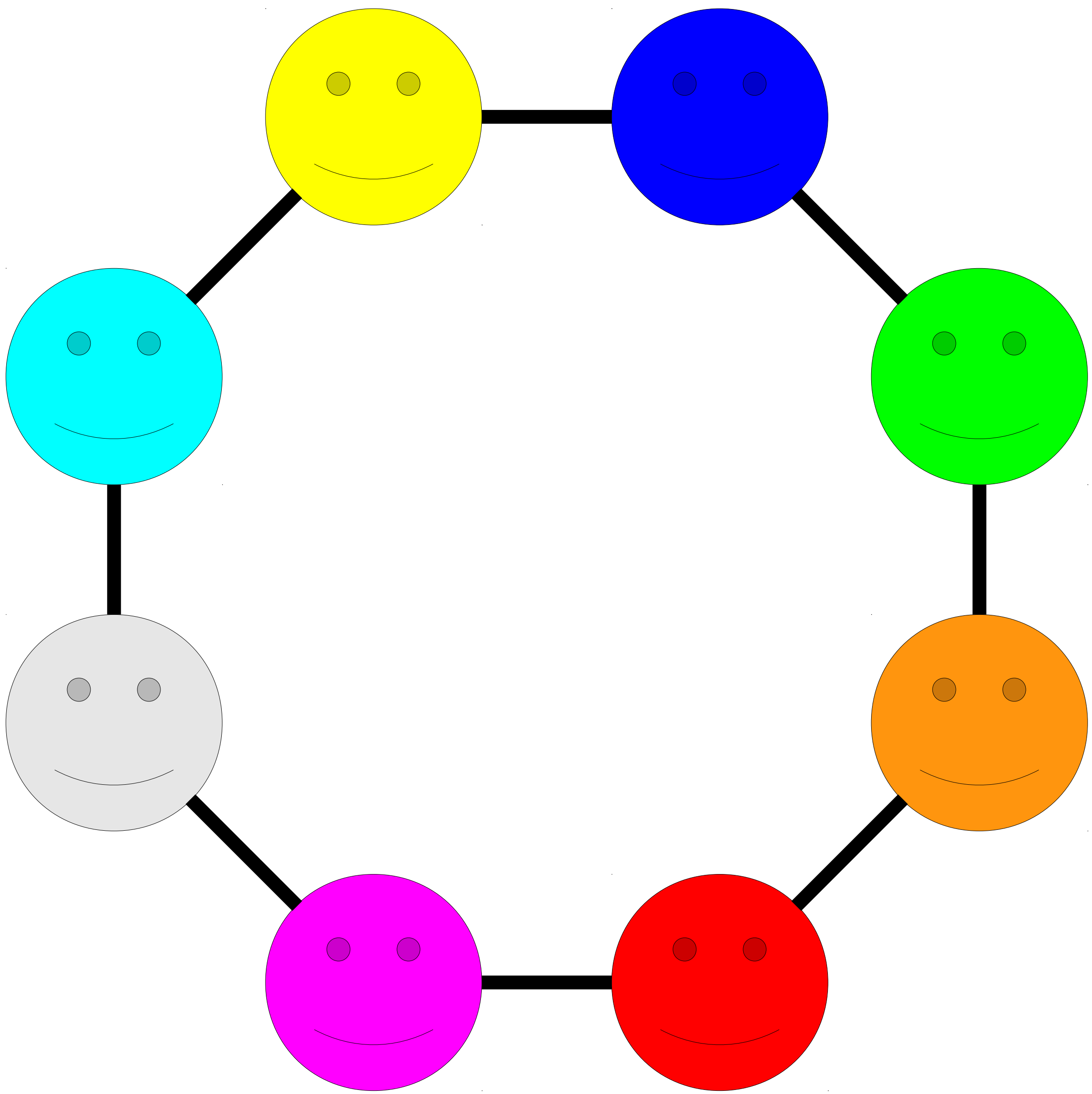
The Wikipedia Graph

- Wikipedia (and the web in general) is a graph!
- Each page is a node.
- There is an edge from one page to another if the first page links to the second.

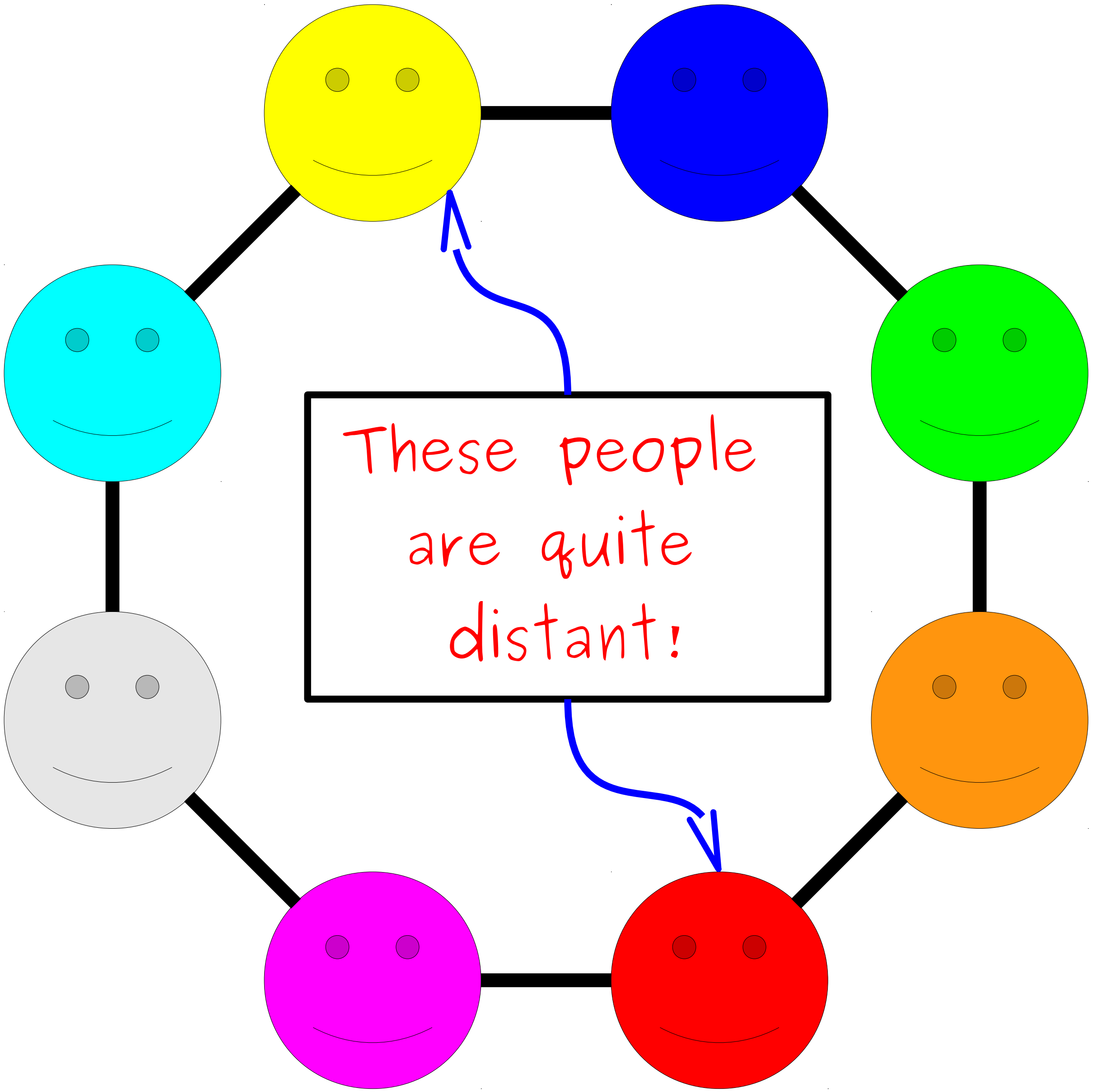
Network Analysis

- We can analyze how nodes in a graph are connected to learn more about the graph.
- How connected are the nodes in the graph?
- How important is each node in the graph?

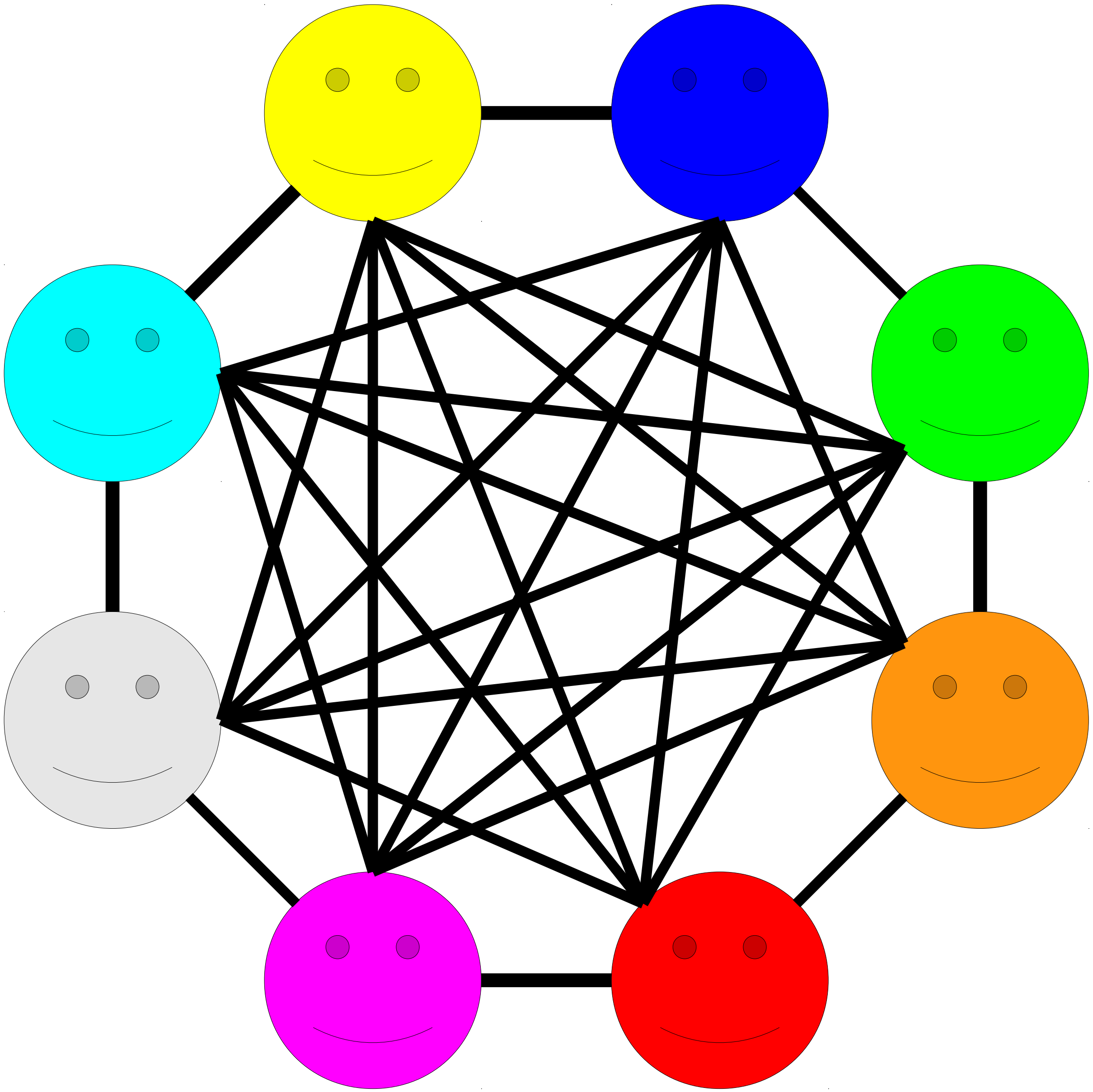
Connectivity



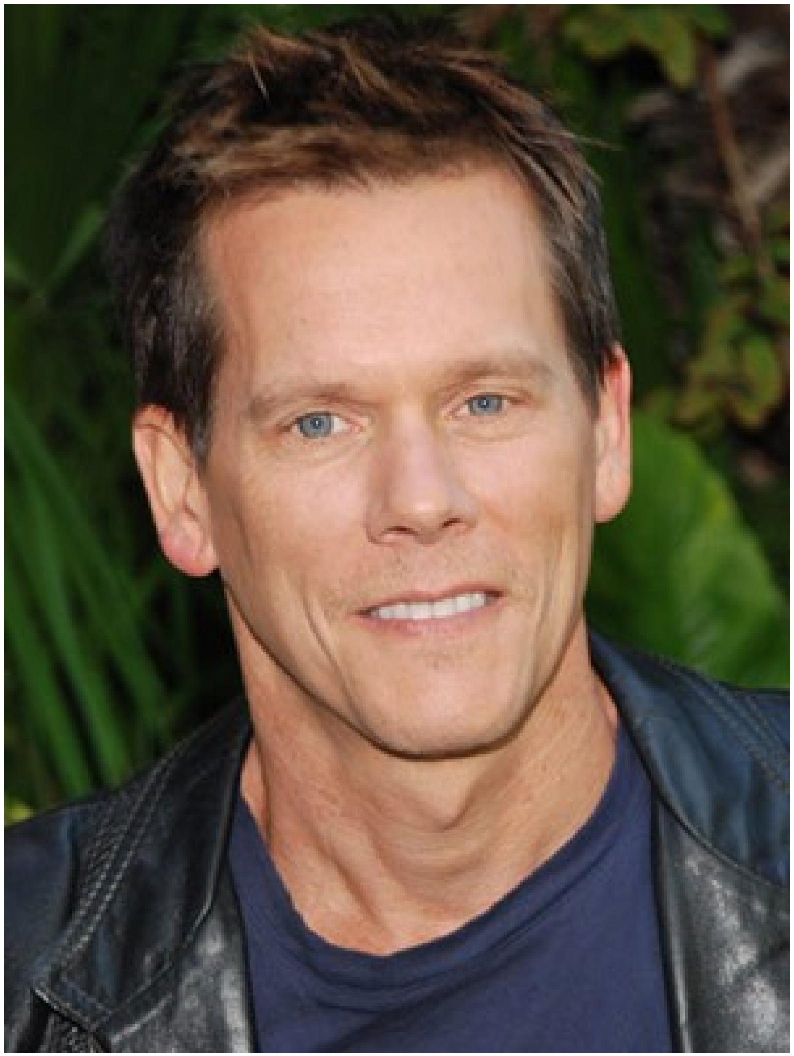
Connectivity



Connectivity



Network Connectivity



- All actors and actresses have a **Bacon number** describing how removed they are from Kevin Bacon.
- Fewer than 1% of all actors and actresses have a Bacon number greater than six.

Image: http://3.bp.blogspot.com/_s59ARuFLHSs/TD-BqVrYtNI/AAAAAAAACx8/5-Ultw06u0s/s1600/Kevinbacon.jpg

Finding Important Nodes

- Suppose that we want to have the computer find “important” articles on Wikipedia.
- We just have the link structure, not the text of the page, the number of edits, the length of the article, etc.
- How might we do this?

Link Analysis

- To find important Wikipedia pages, let's look at the links between pages.
- We'll make two assumptions:
  - The more important an article is, the more pages will link to it.
  - The more important an article is, the more that its links matter.
- **An article is important if other important articles link to it.**

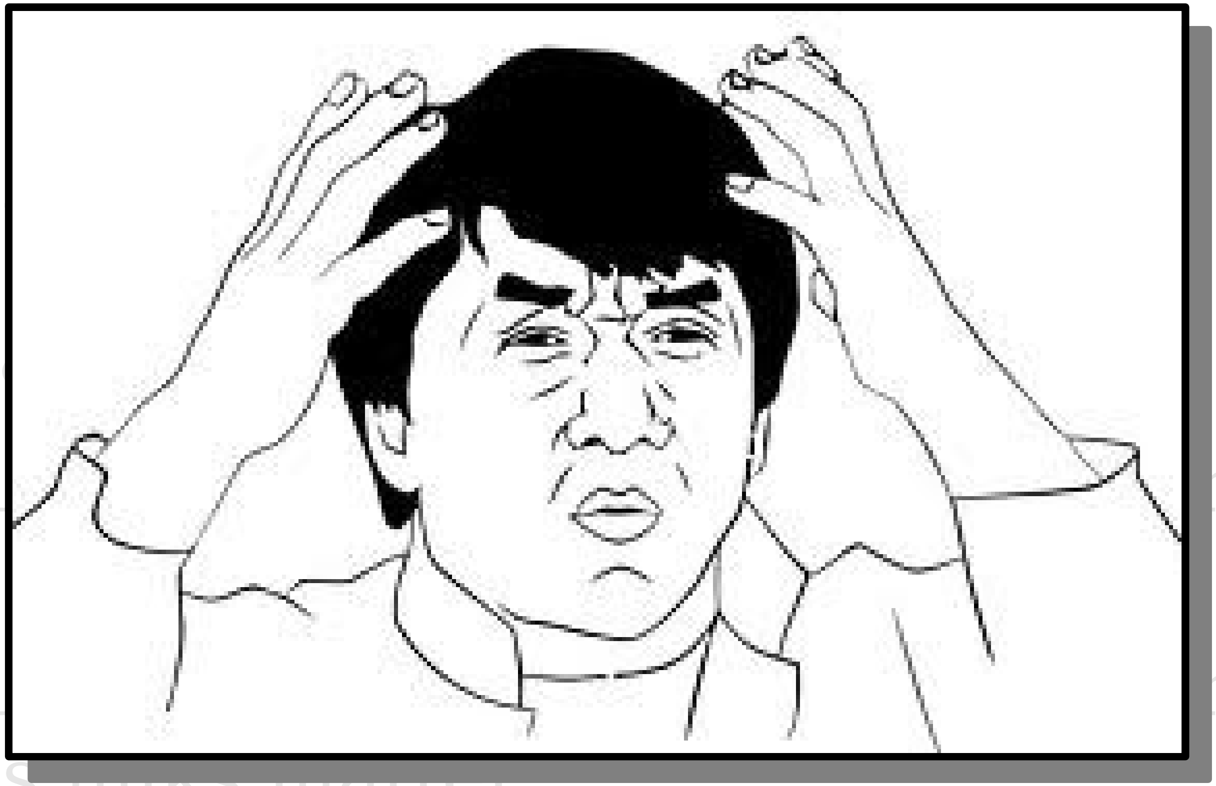
Link Analysis

To find ... let's
look at ...

We'll m...

The m... more
pages ...

The m... more
that its links matter.



- **An article is important if other important articles link to it.**

The Random Surfer Model

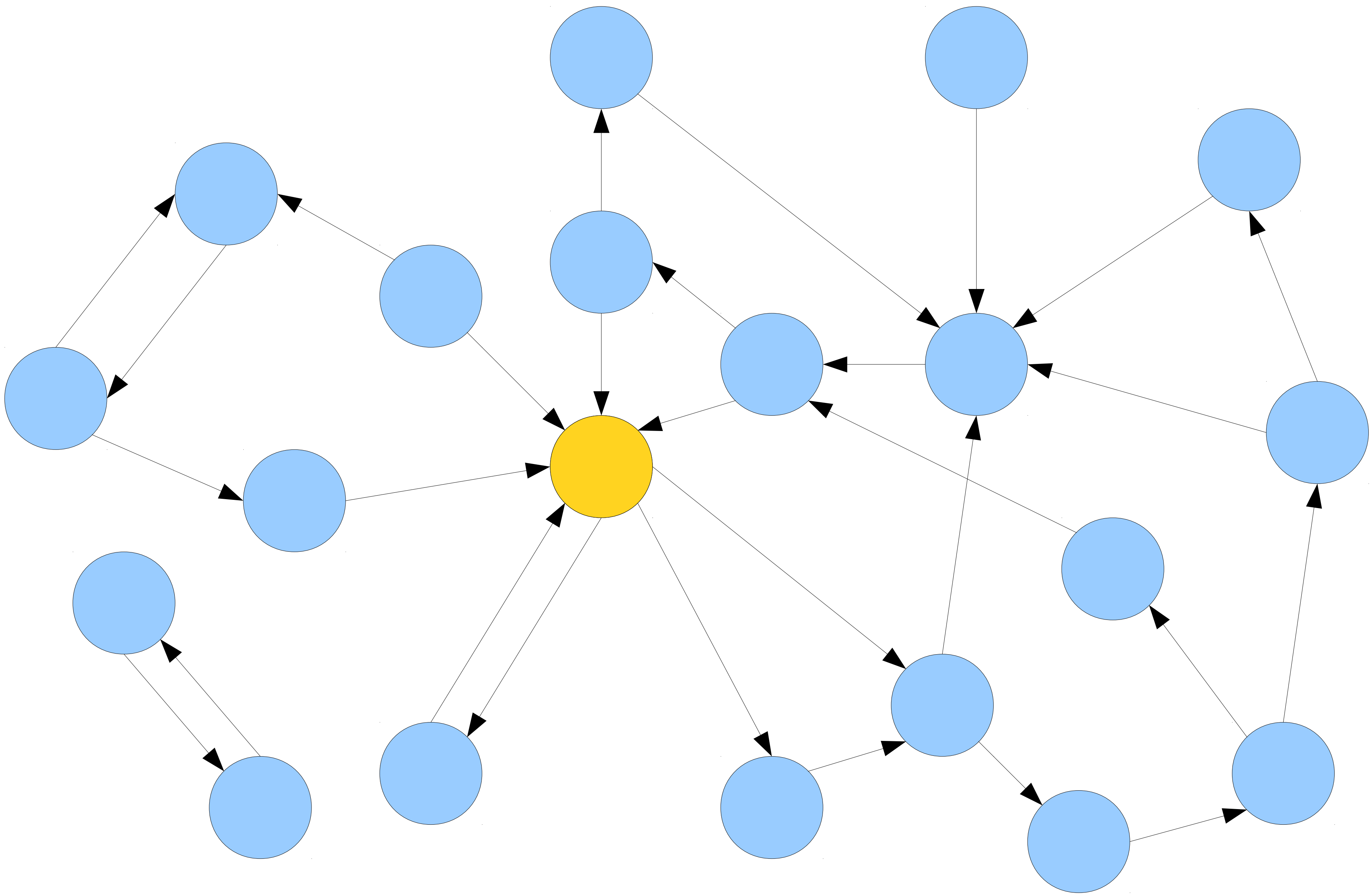
http://farm1.static.flickr.com/192/492079742_30a94de366.jpg

(seriously though)

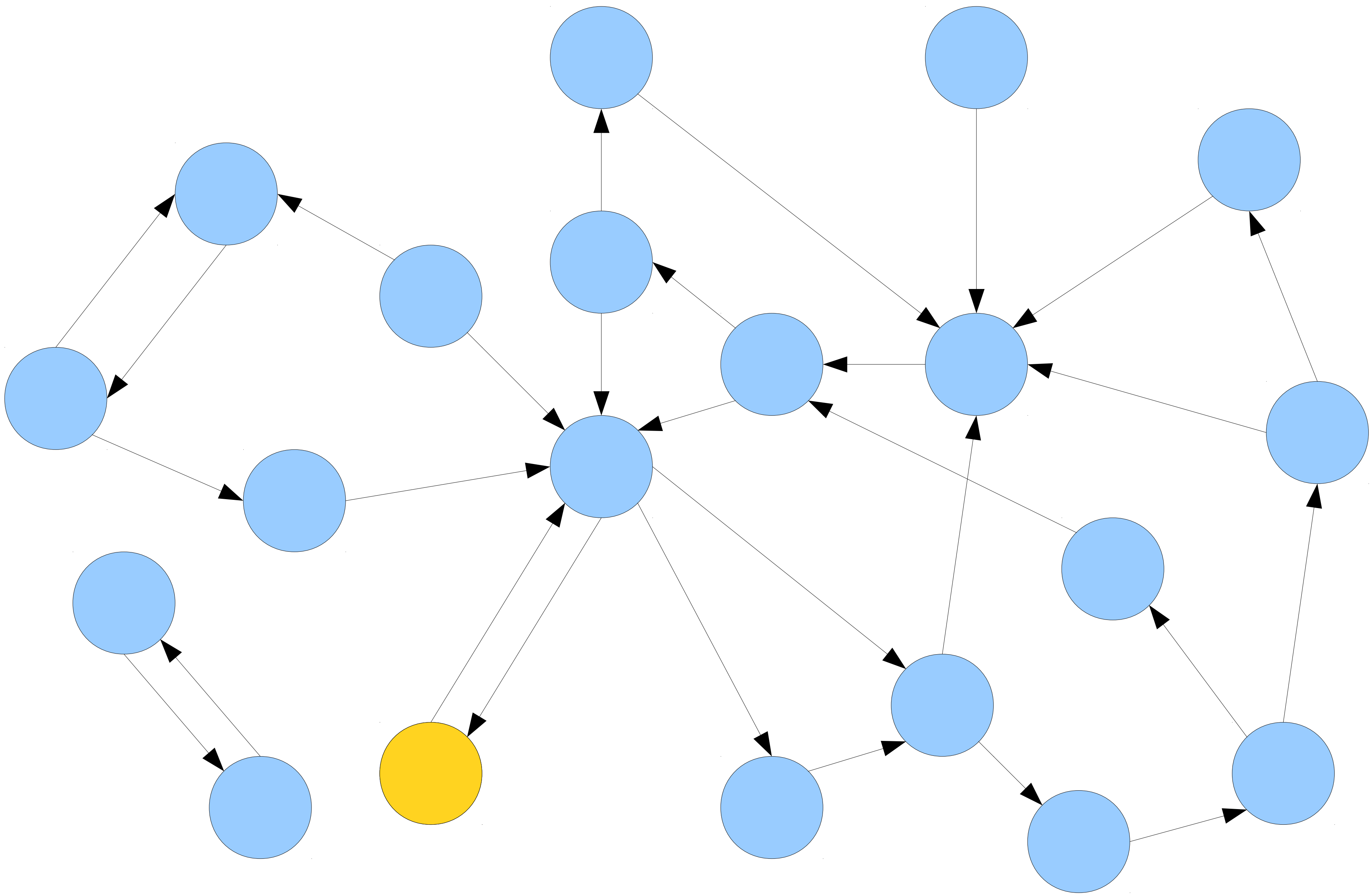
The Random Surfer Model

- Think about the behavior of a Wikipedia reader who randomly surfs Wikipedia.
- Visits some initial page at random.
- From there, the user either
  - Clicks a random link on the page to some other article, or
  - hits the “random page” link to visit a totally random page.

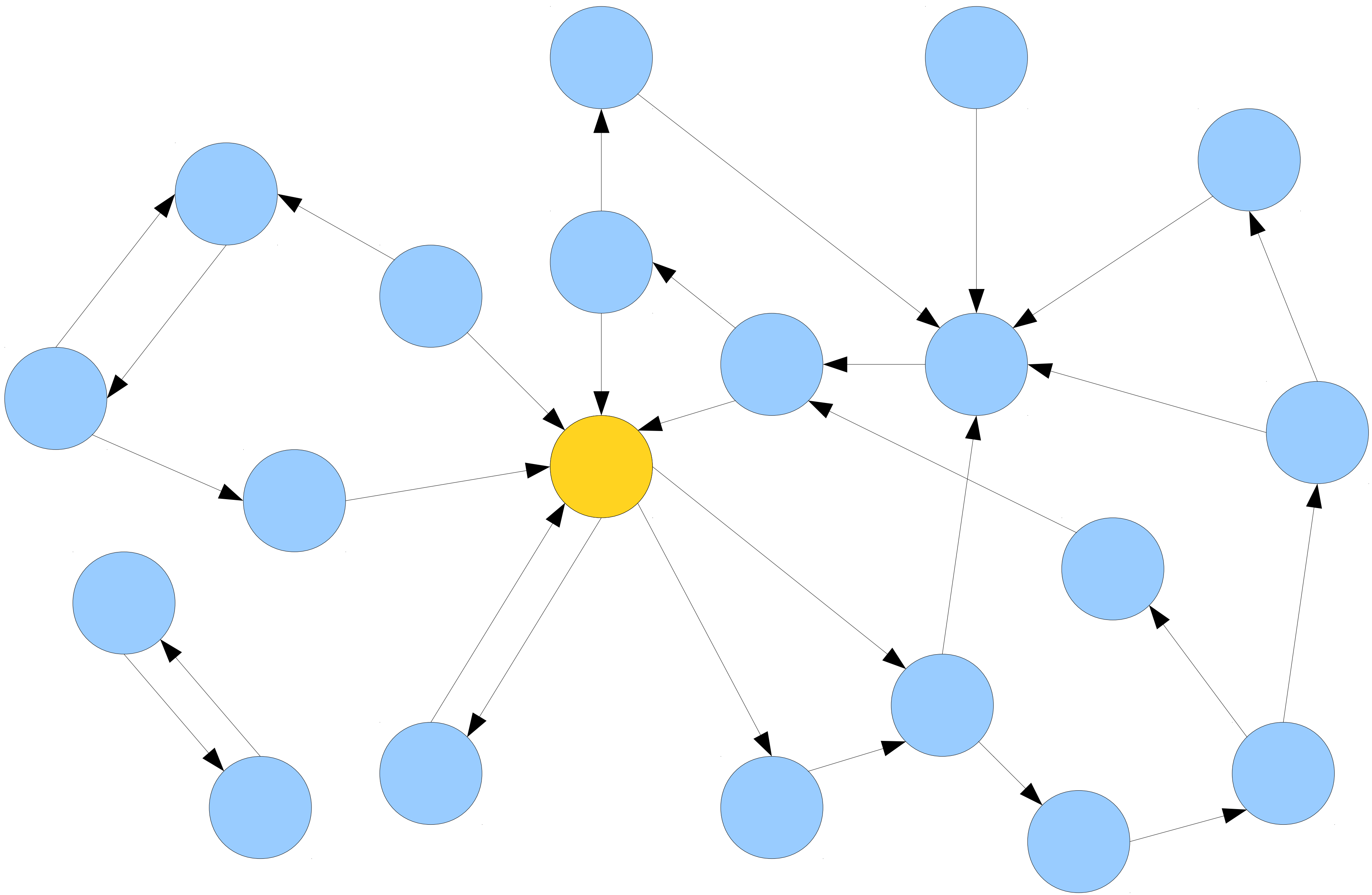
The Random Surfer Model



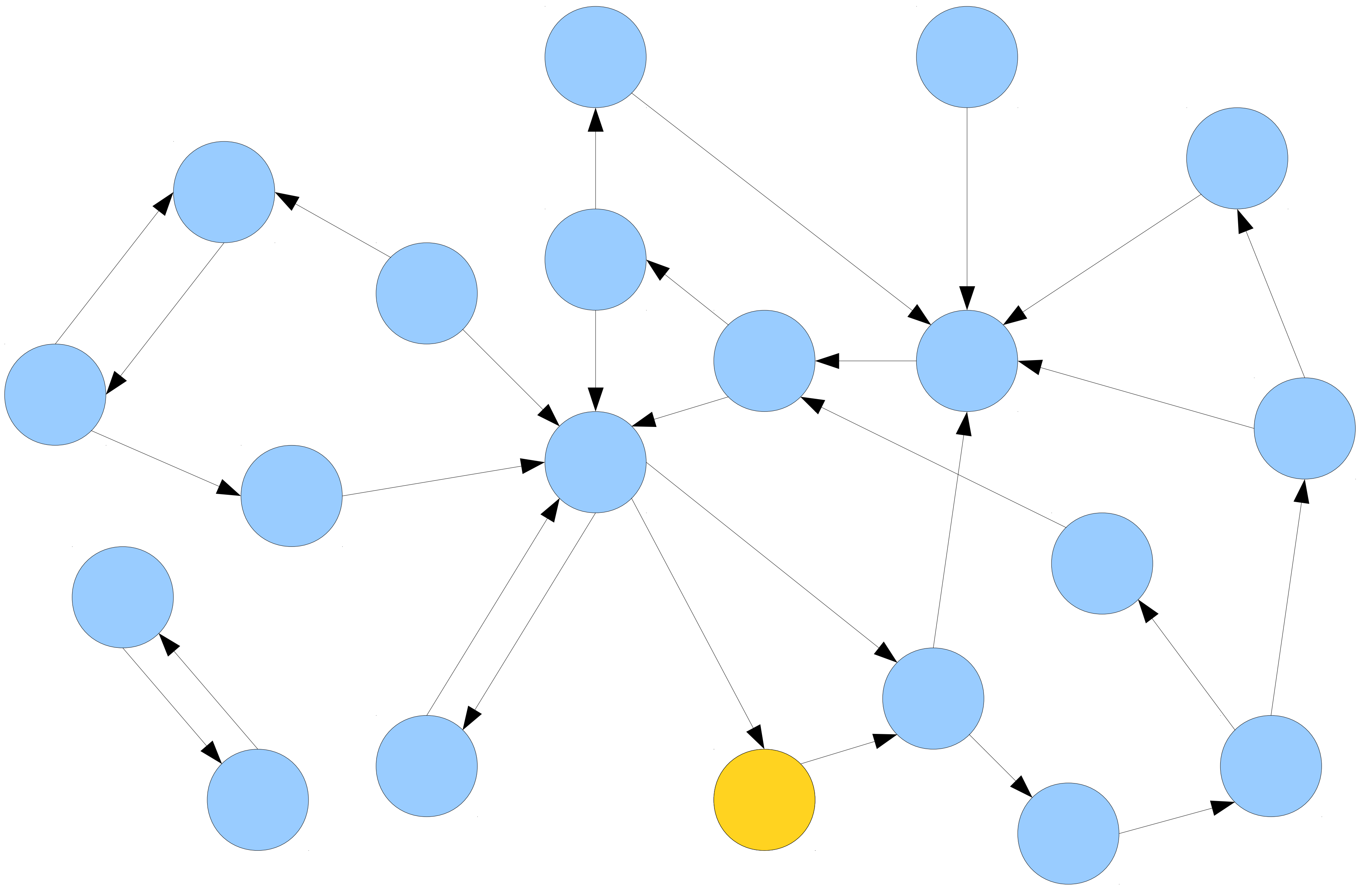
The Random Surfer Model



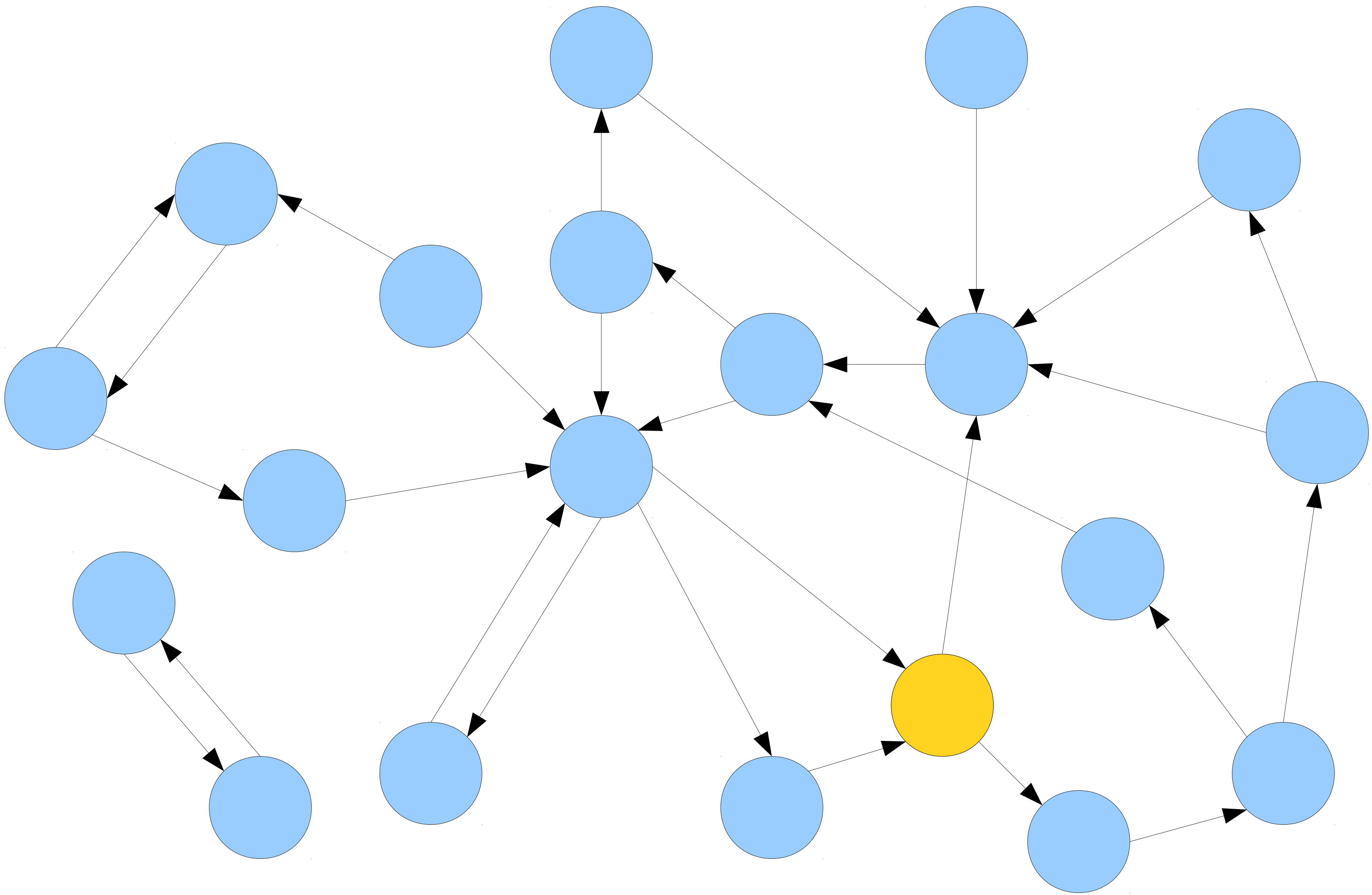
The Random Surfer Model



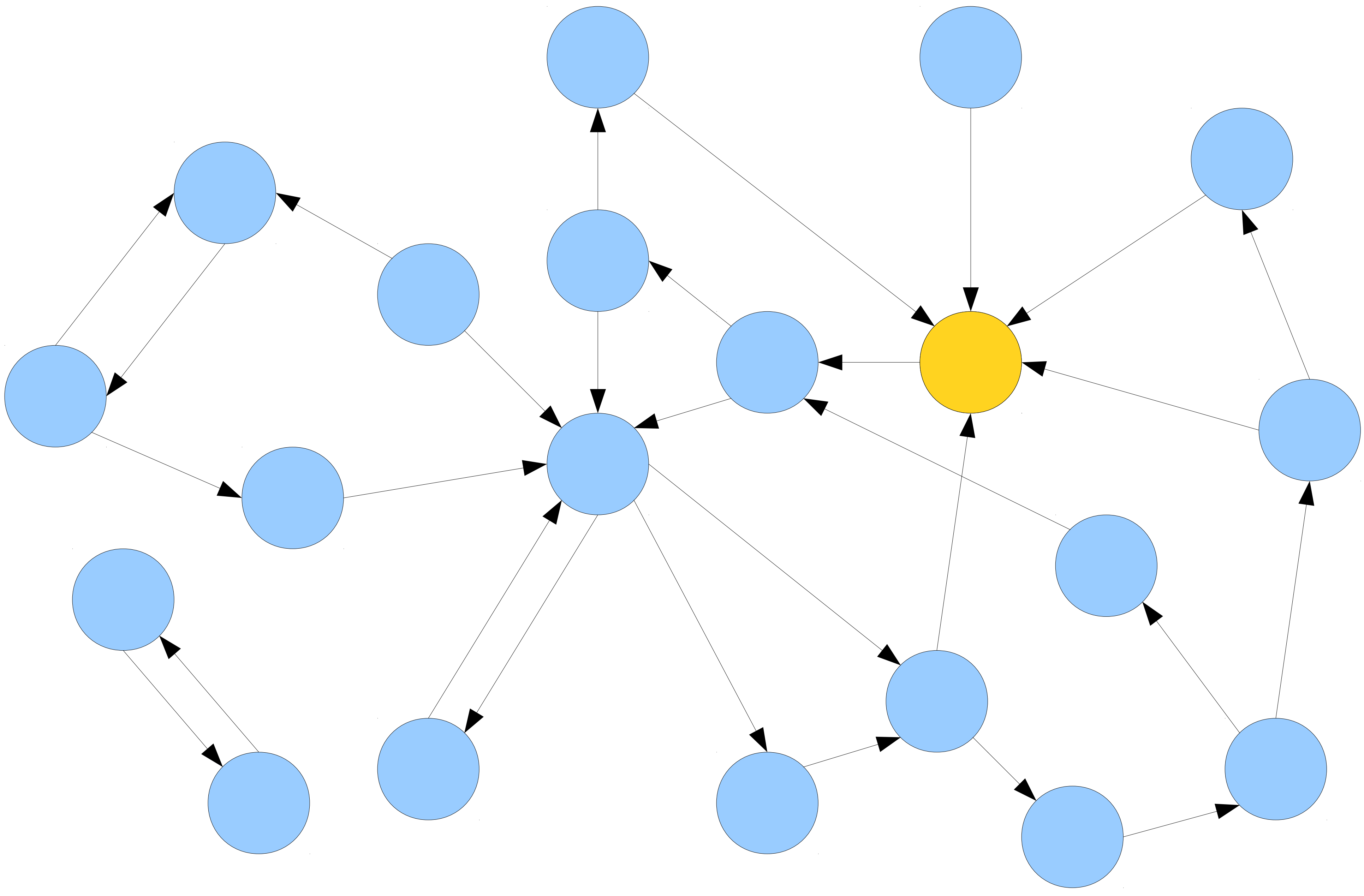
The Random Surfer Model



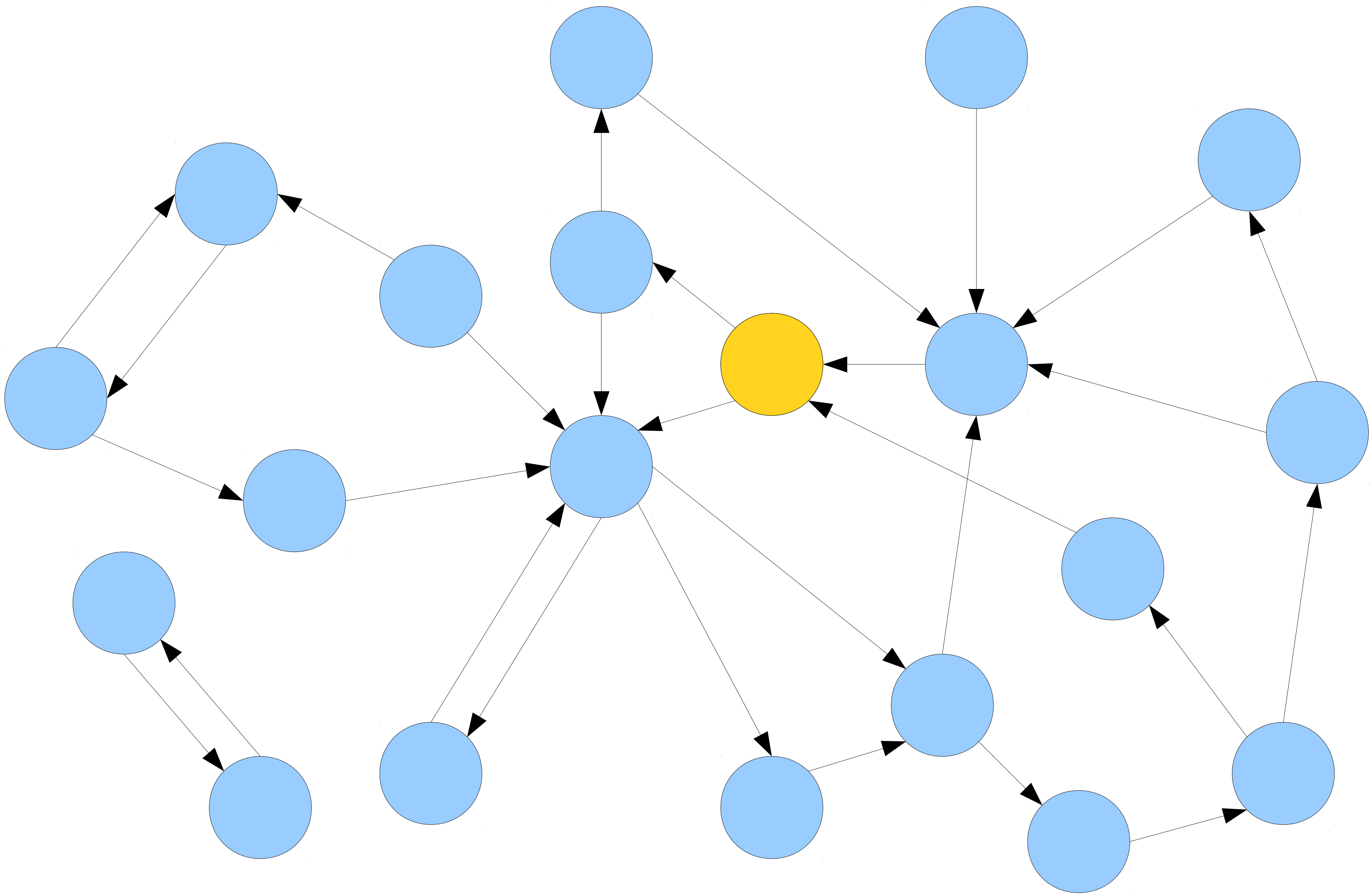
The Random Surfer Model



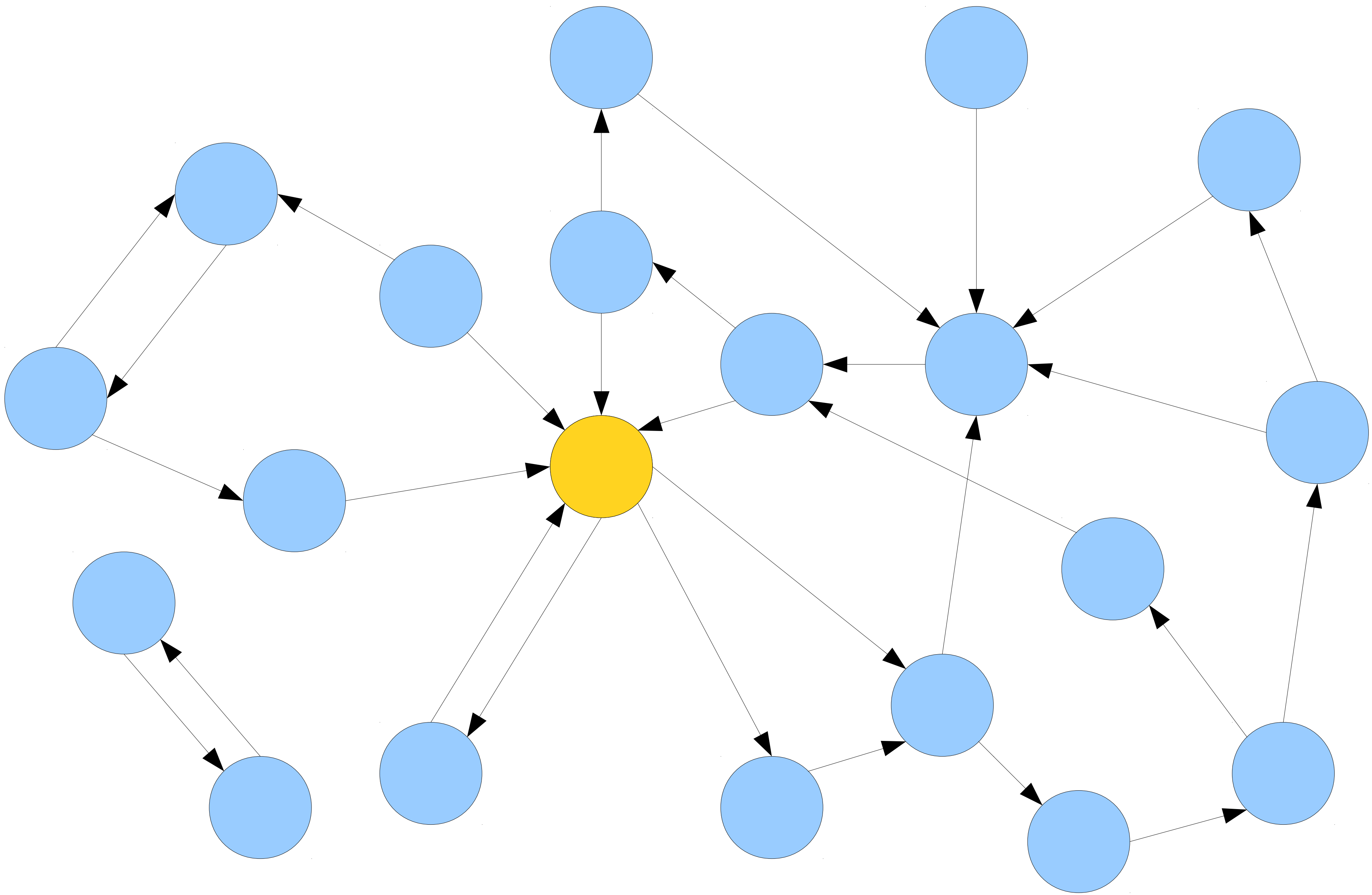
The Random Surfer Model



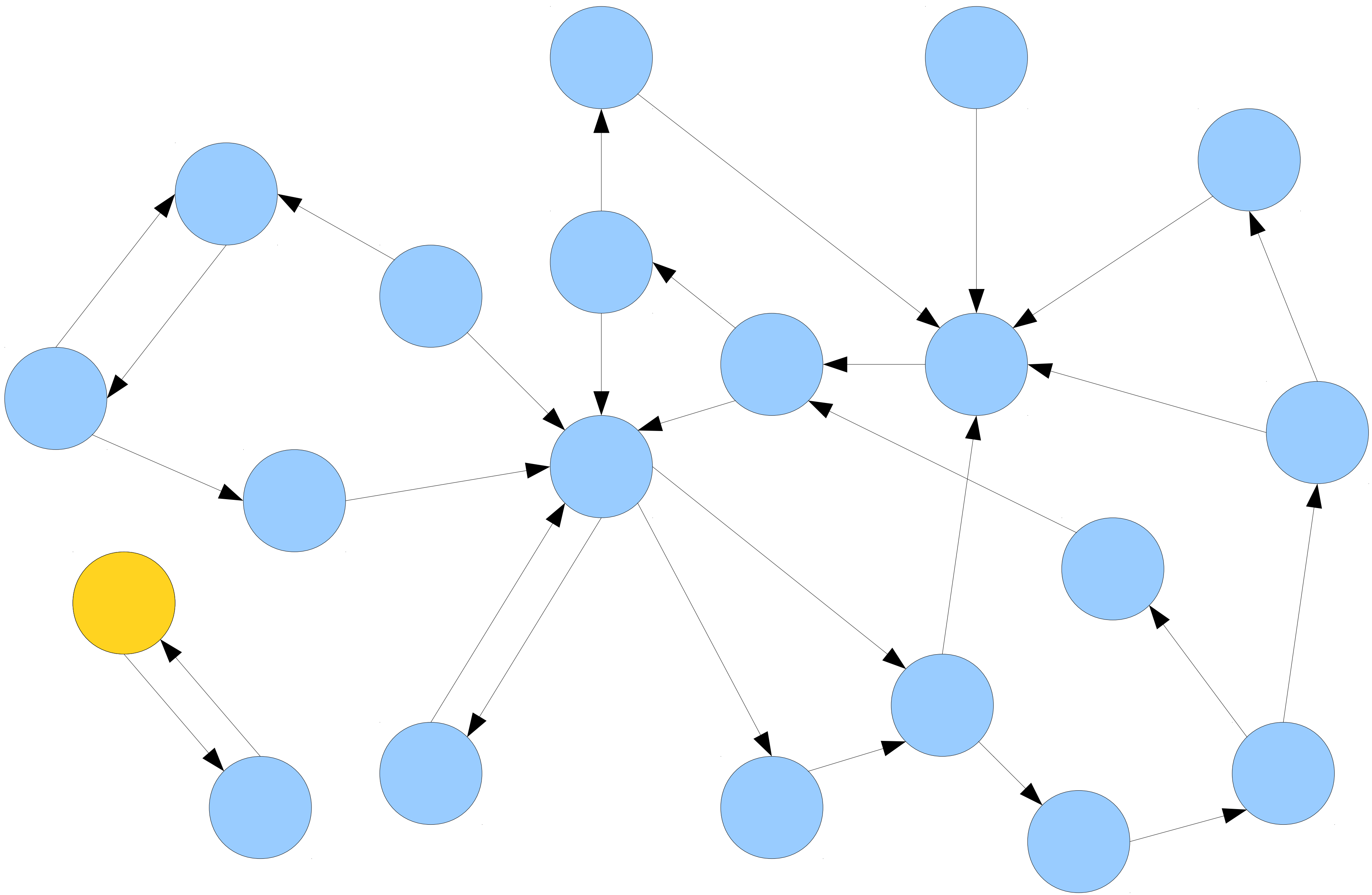
The Random Surfer Model



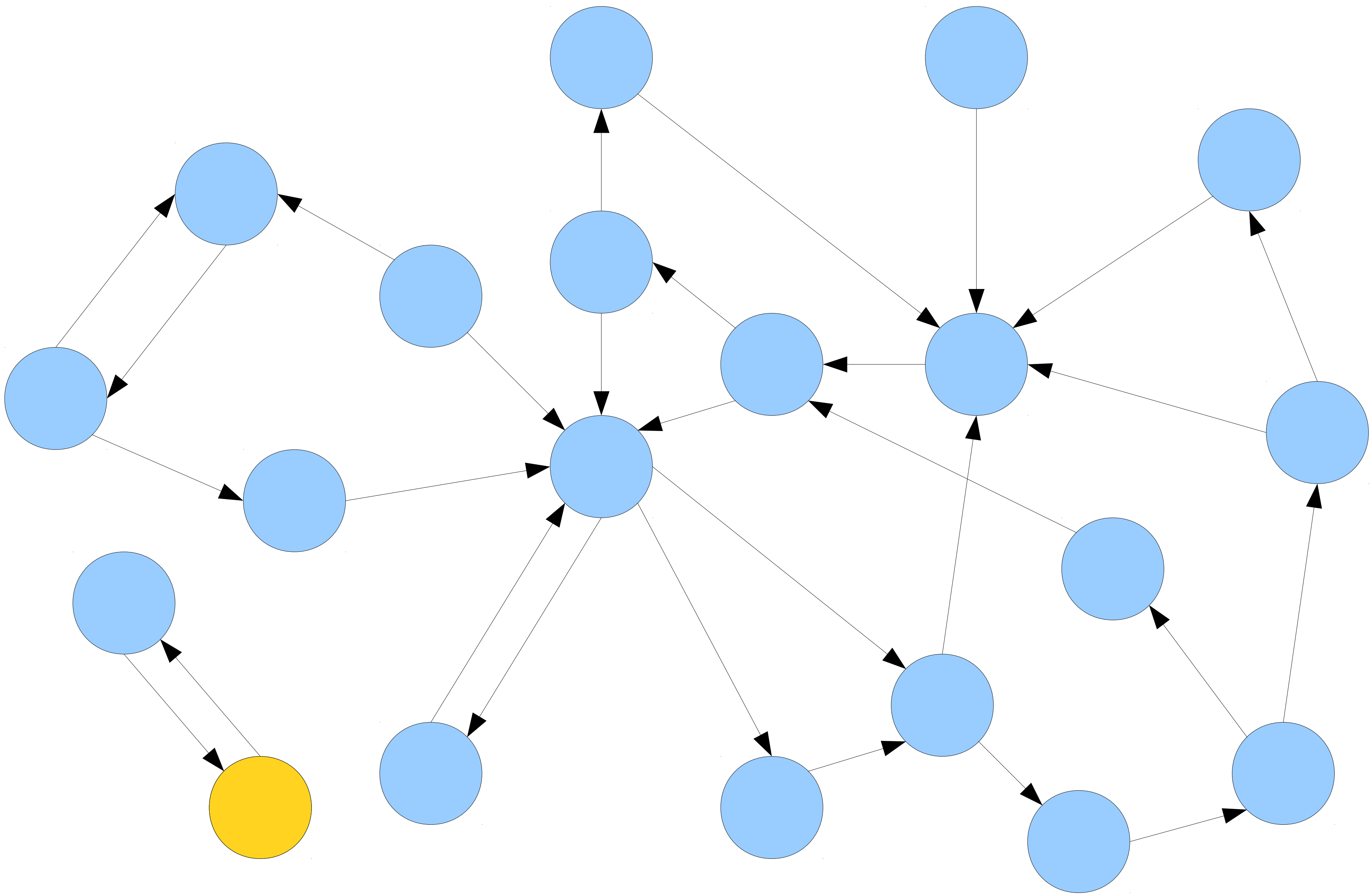
The Random Surfer Model



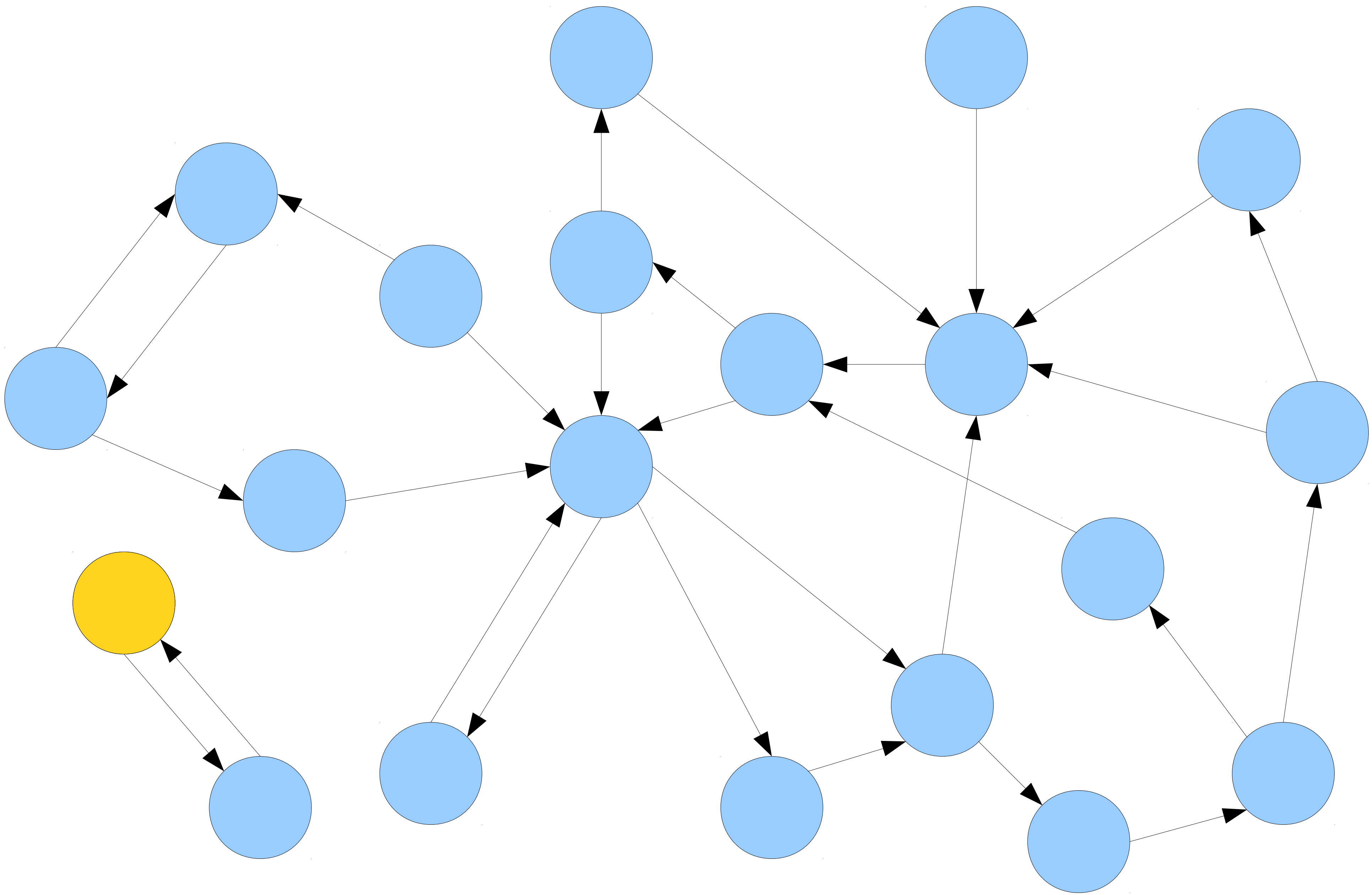
The Random Surfer Model



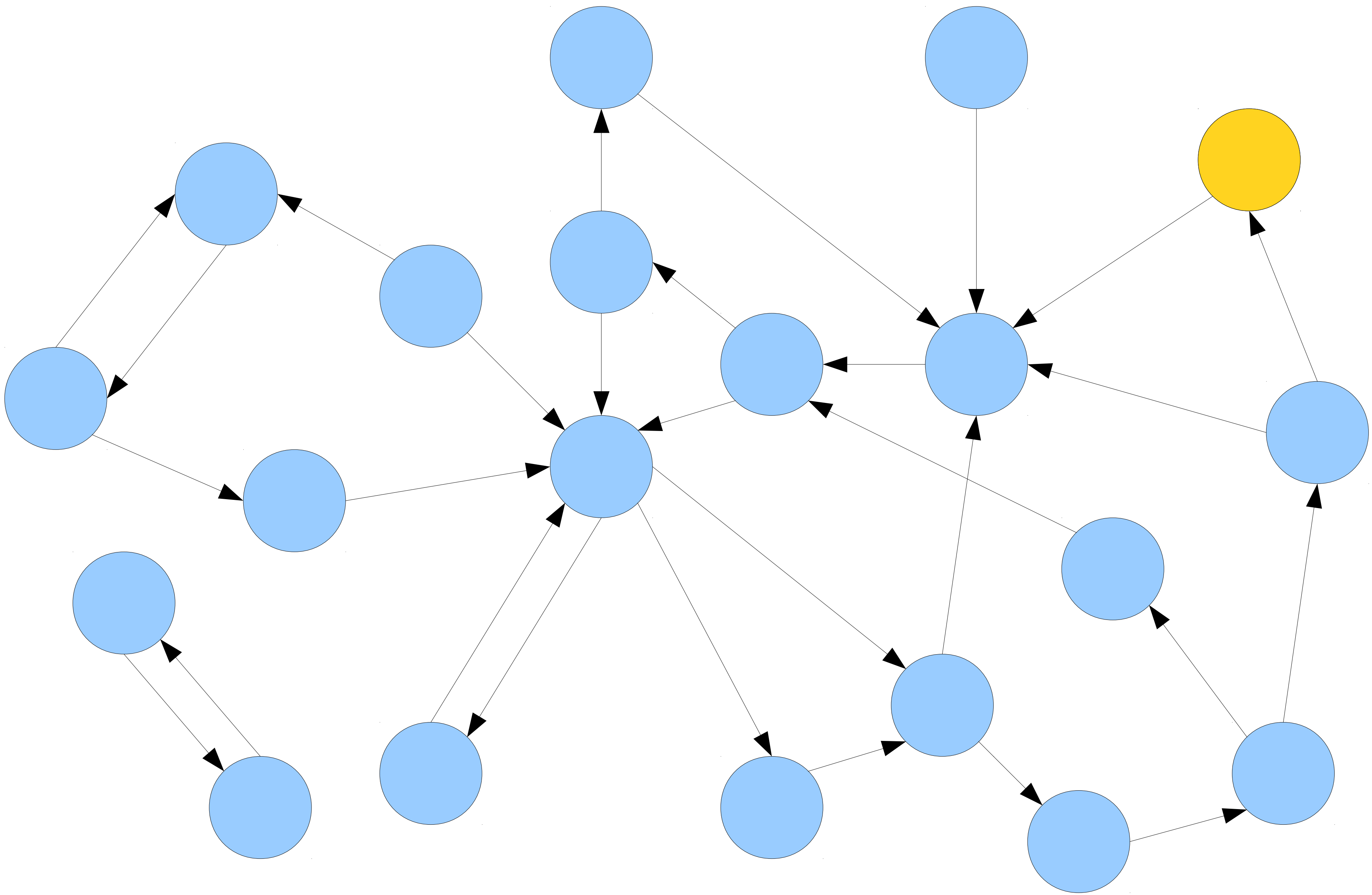
The Random Surfer Model



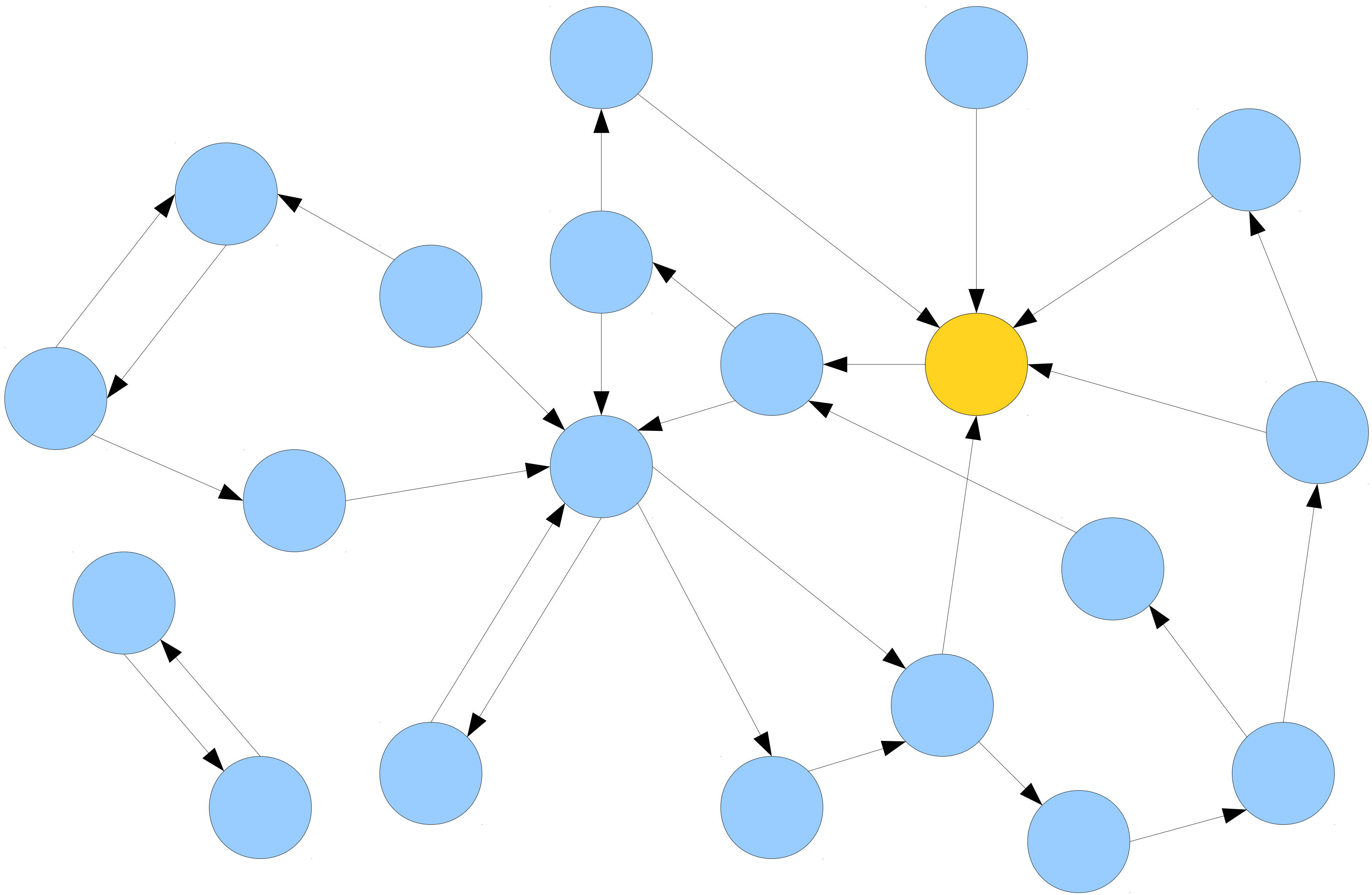
The Random Surfer Model



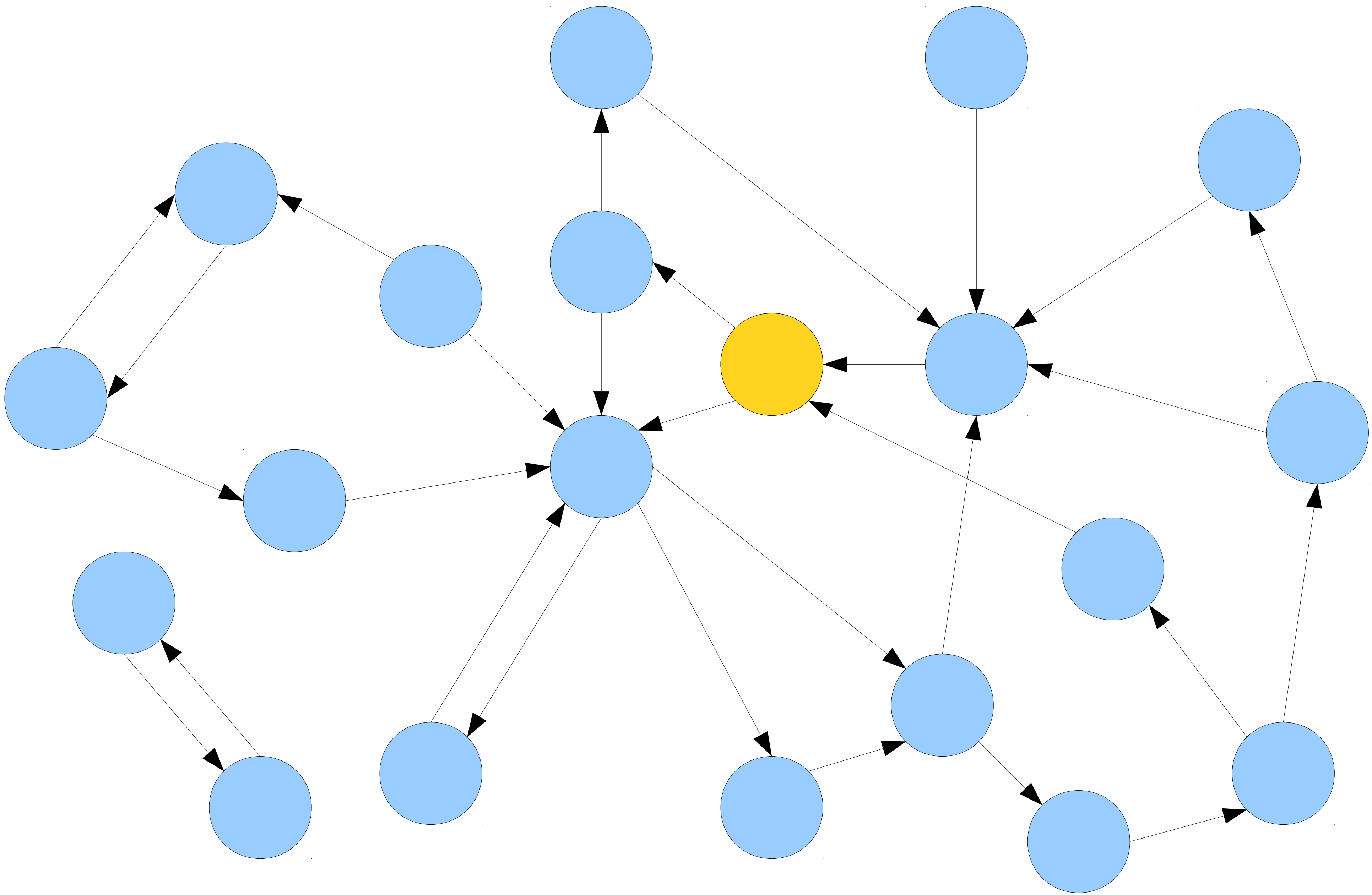
The Random Surfer Model



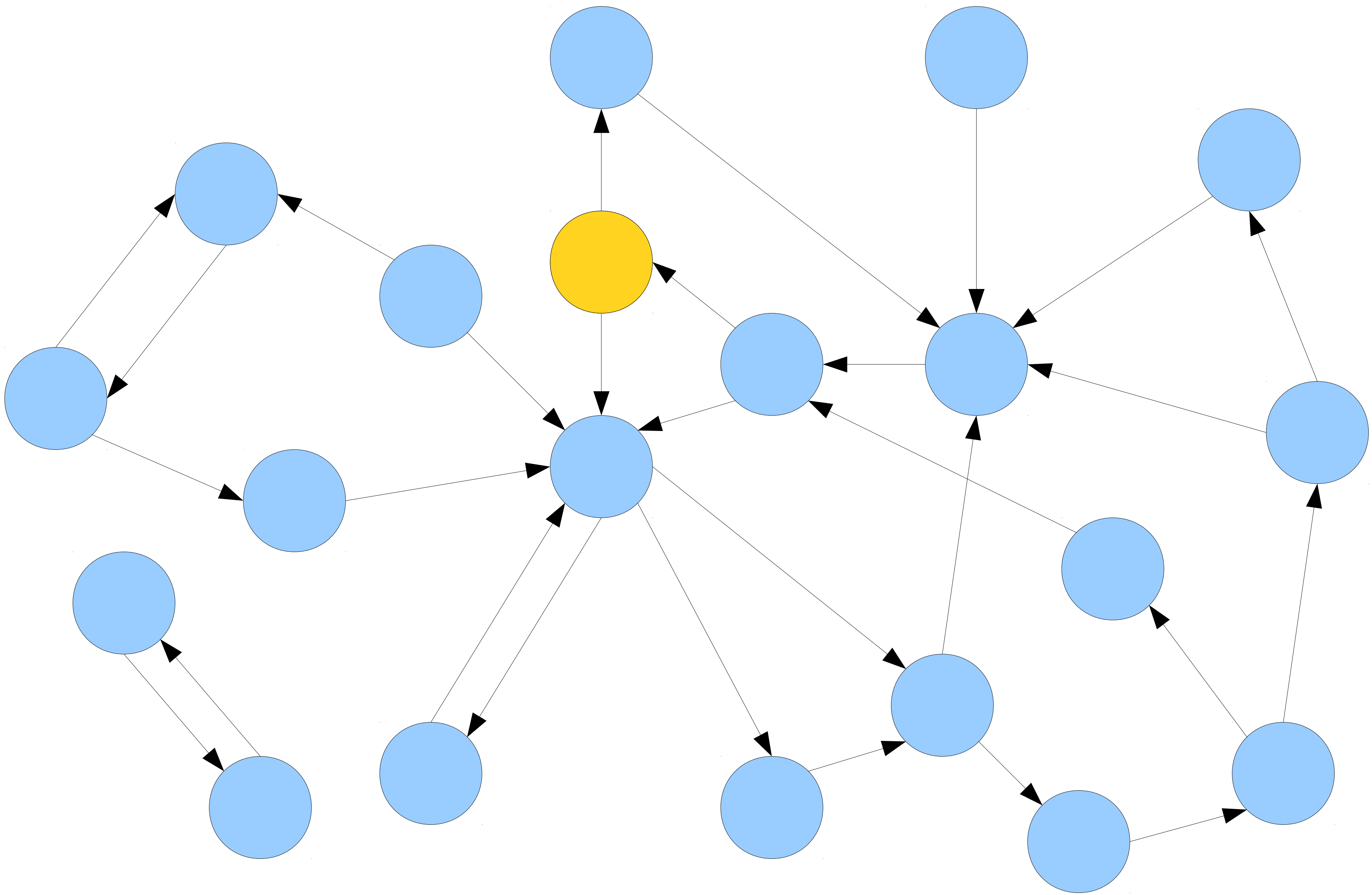
The Random Surfer Model



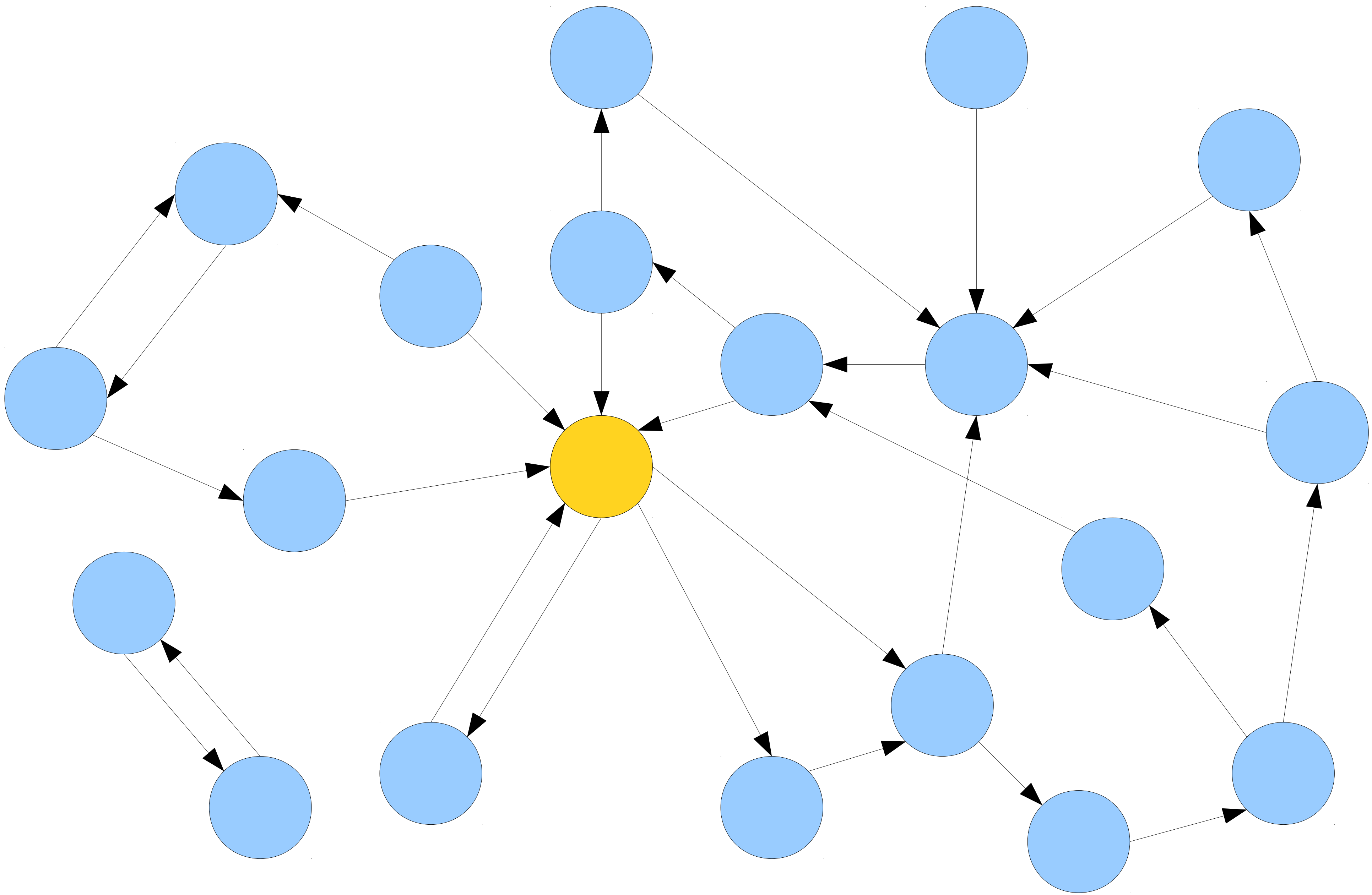
The Random Surfer Model



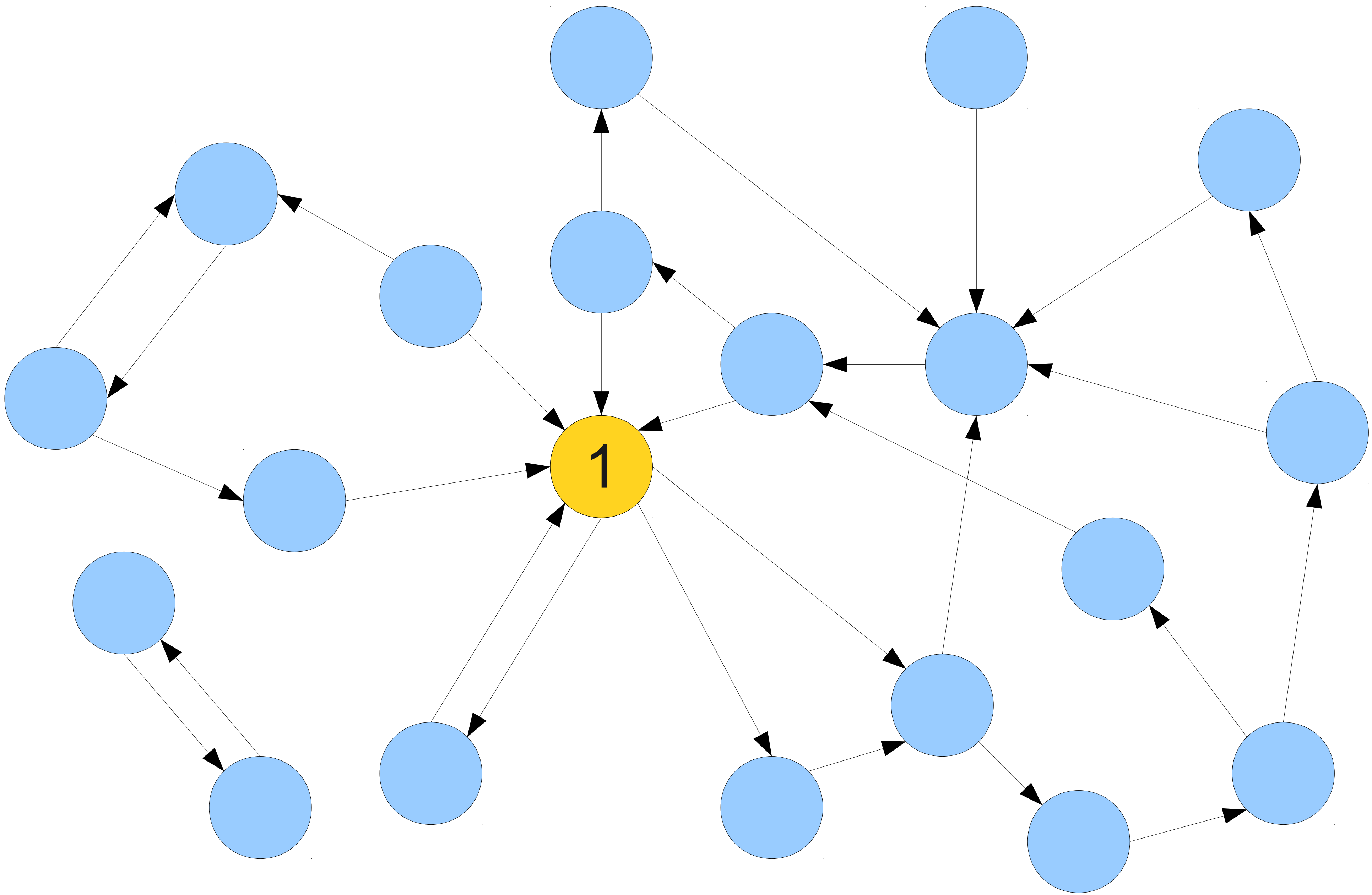
The Random Surfer Model



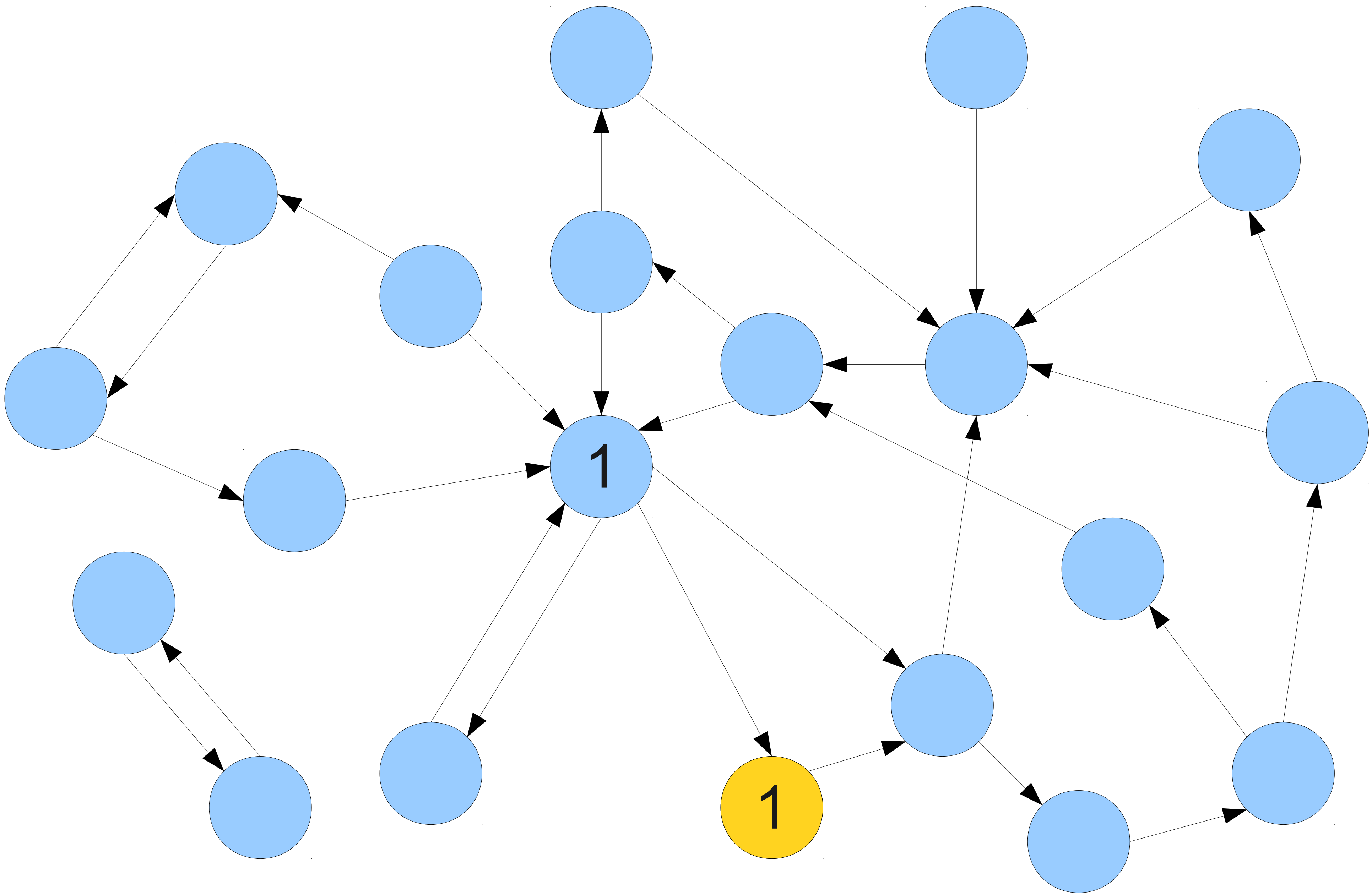
The Random Surfer Model



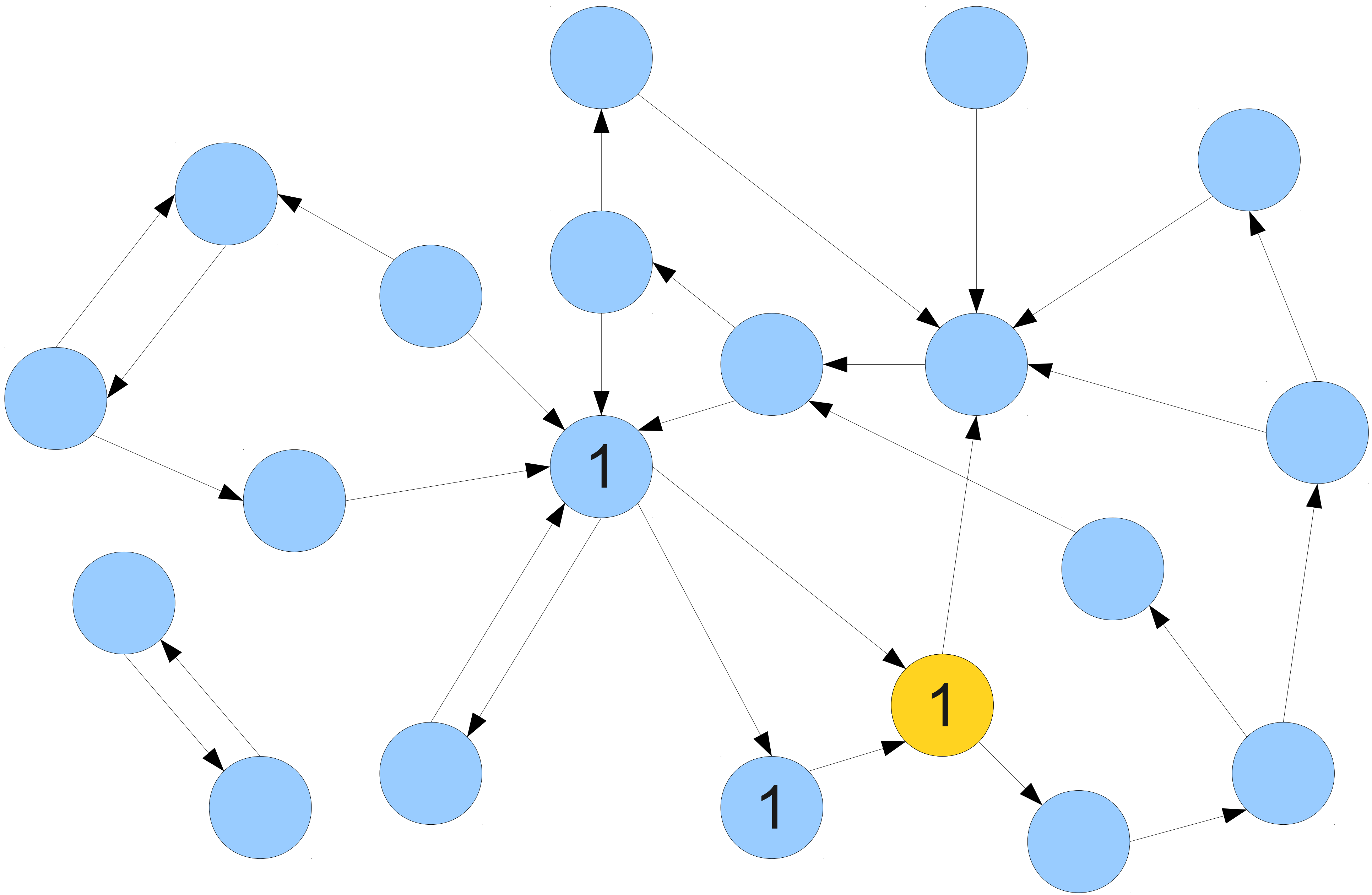
The Random Surfer Model



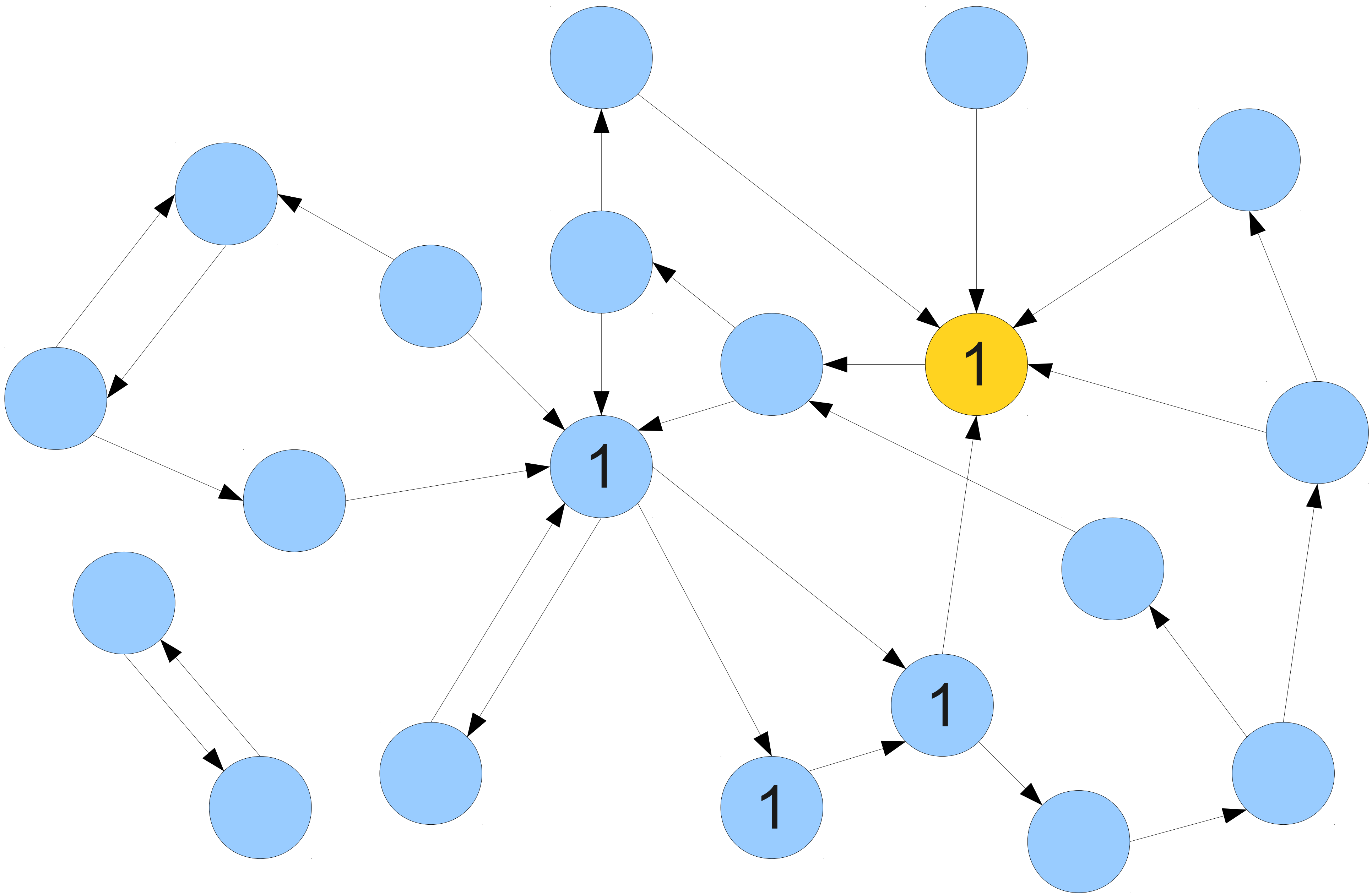
The Random Surfer Model



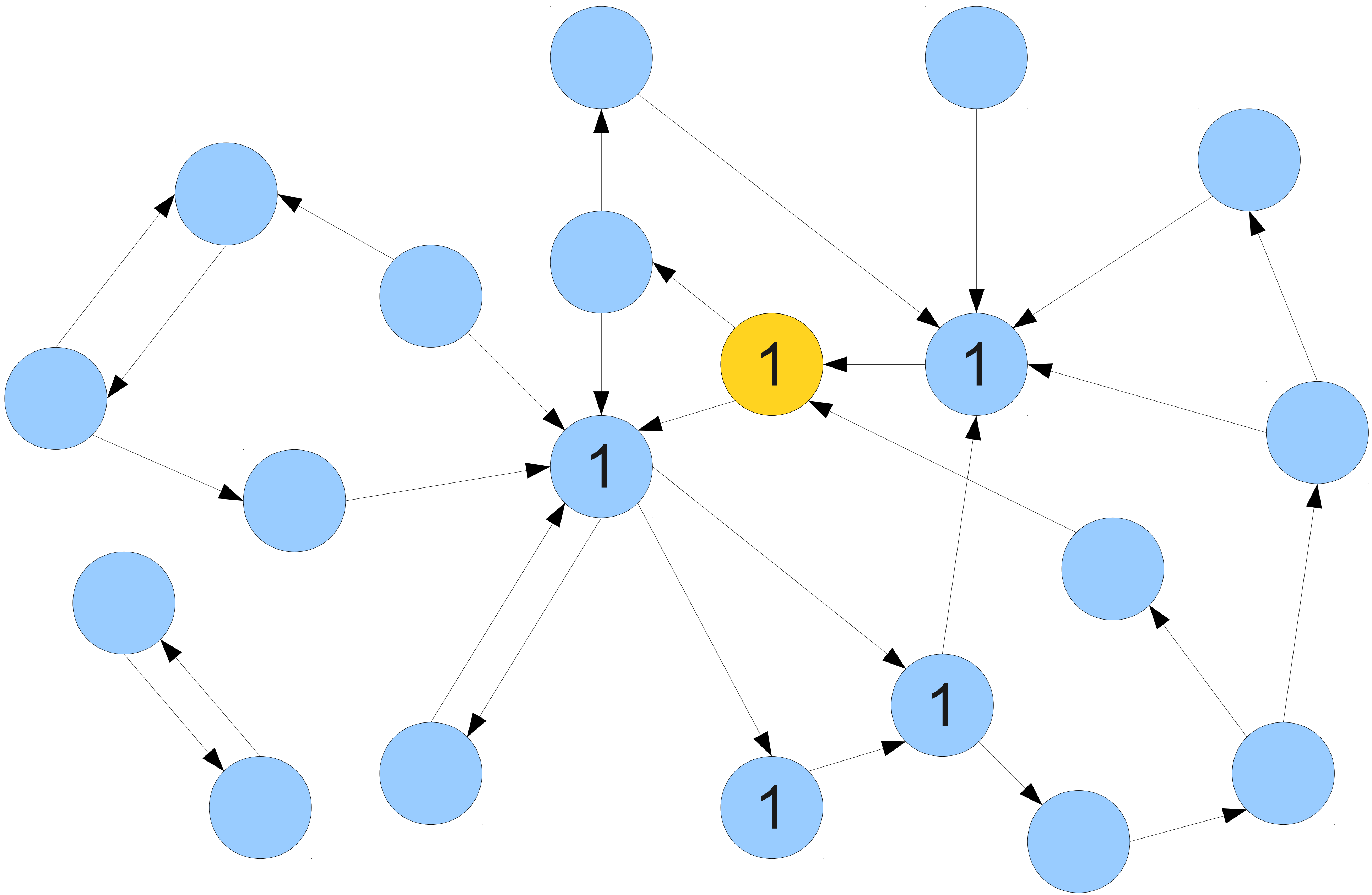
The Random Surfer Model



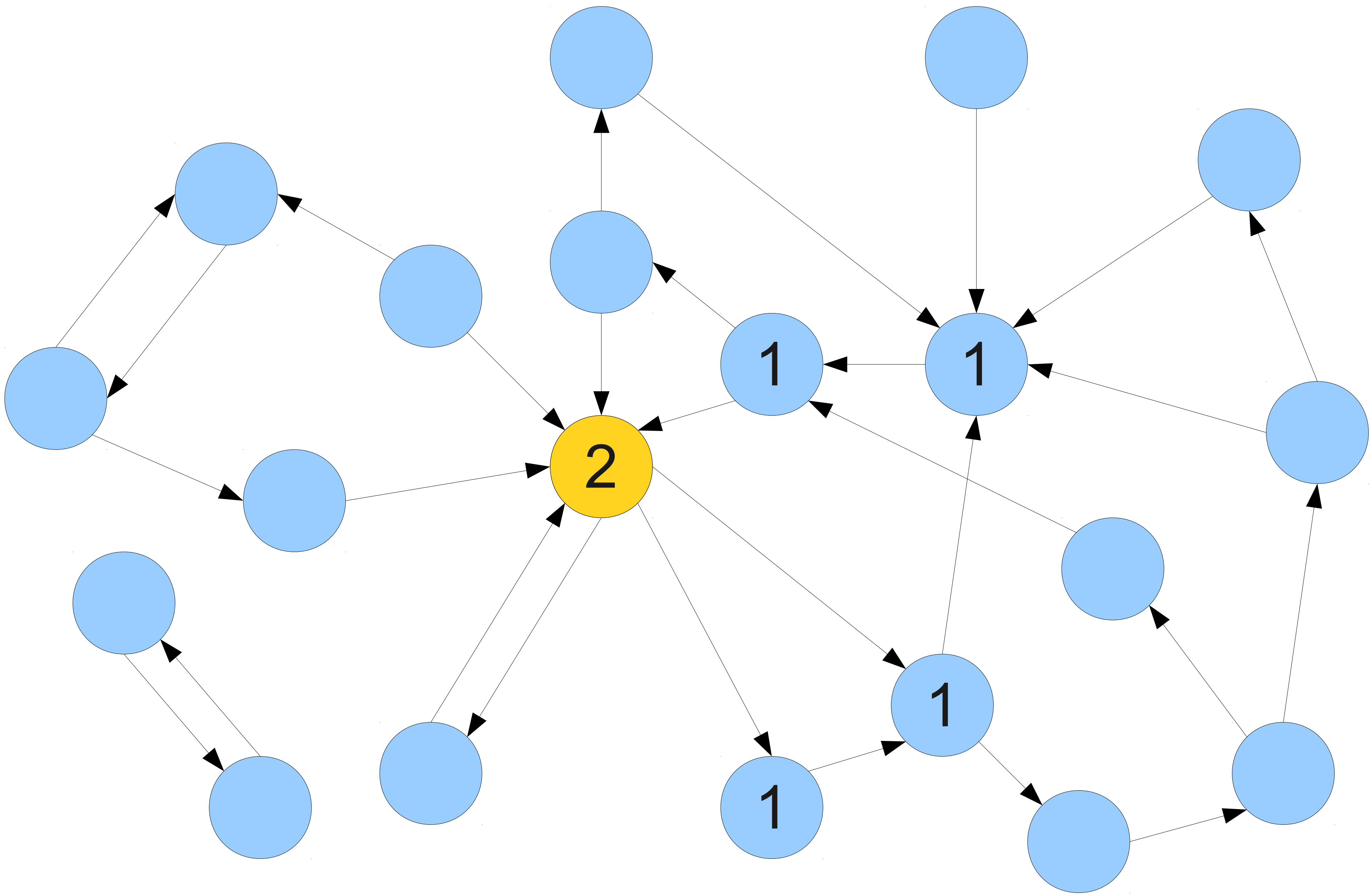
The Random Surfer Model



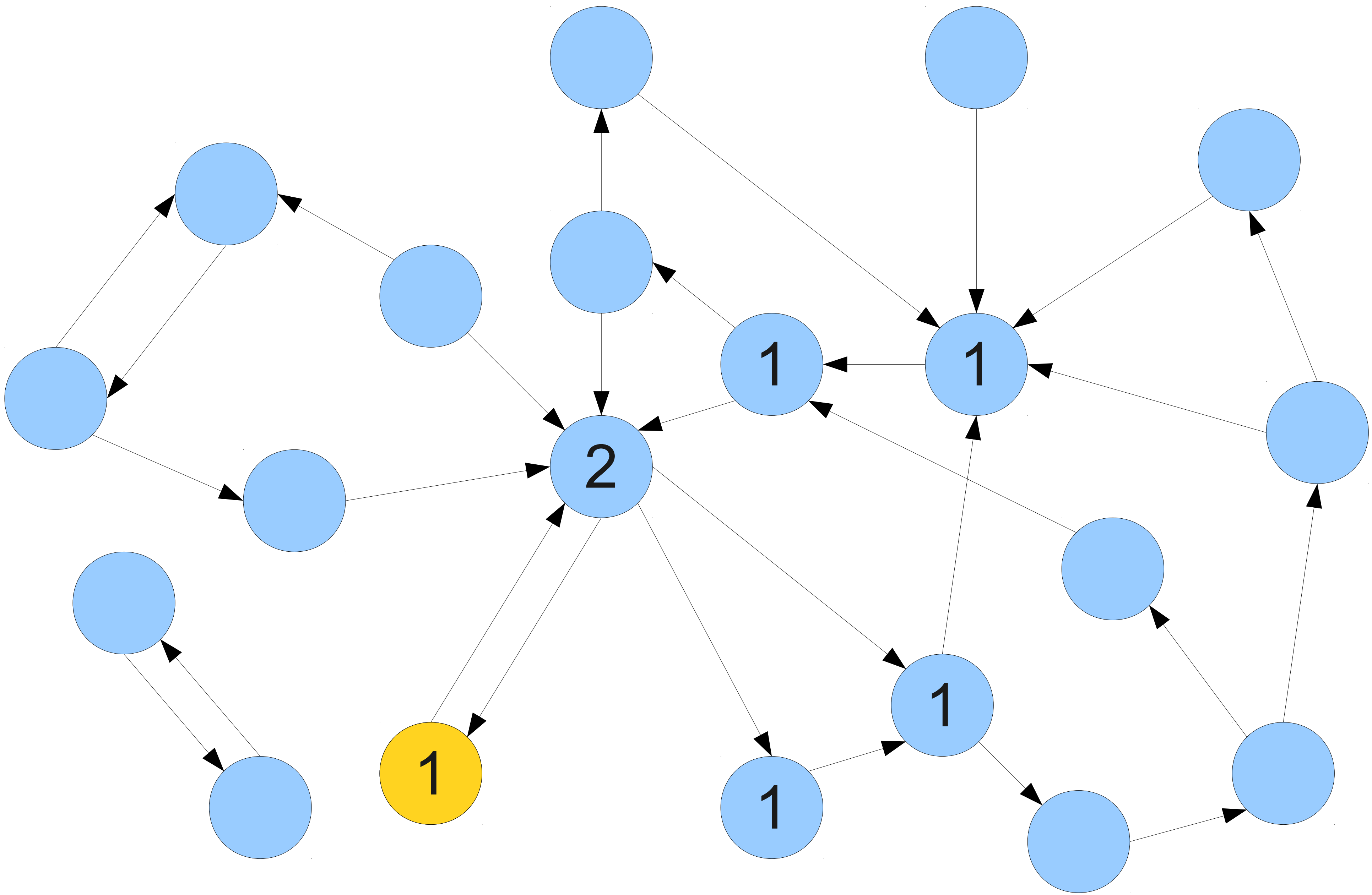
The Random Surfer Model



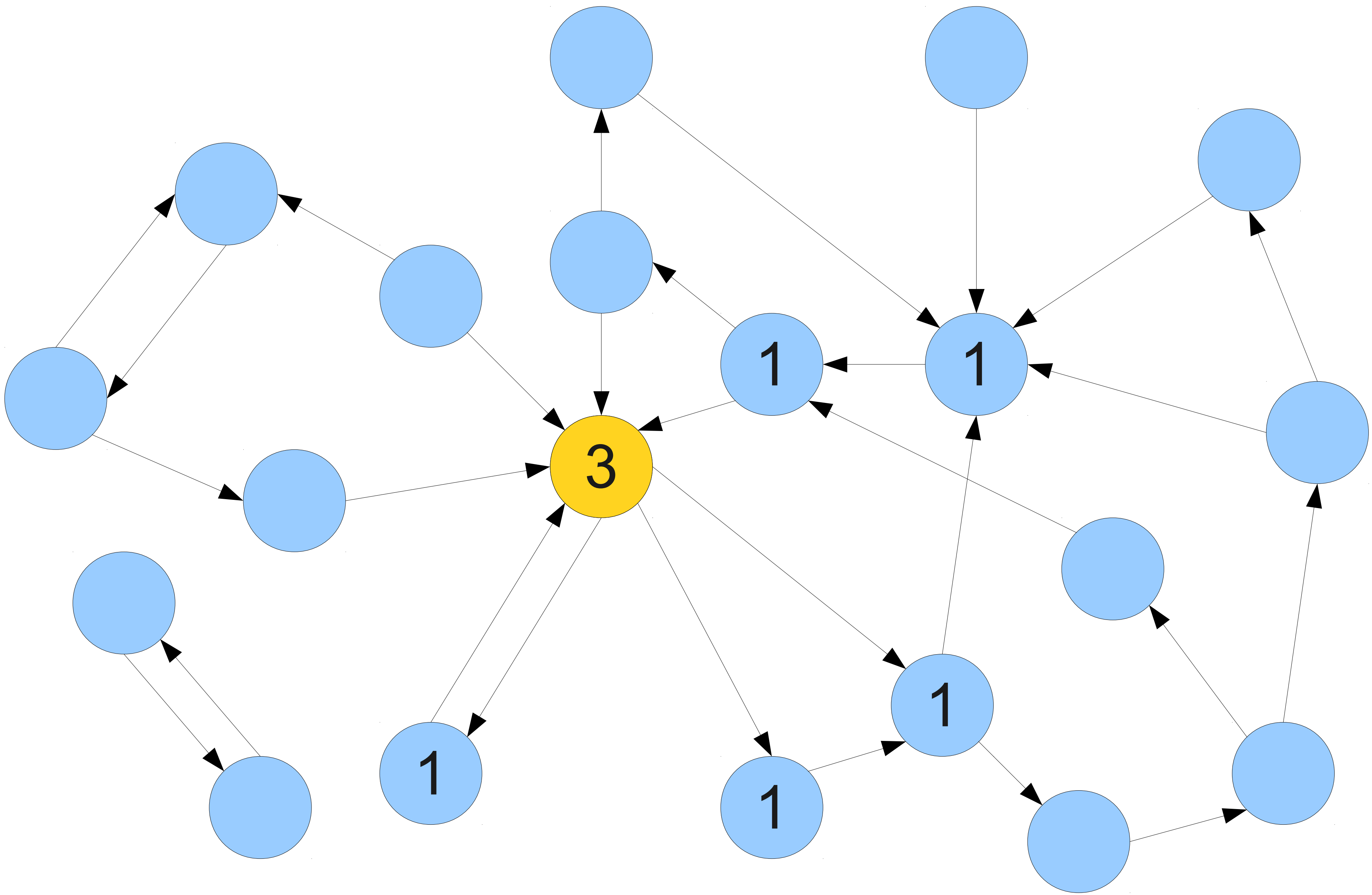
The Random Surfer Model



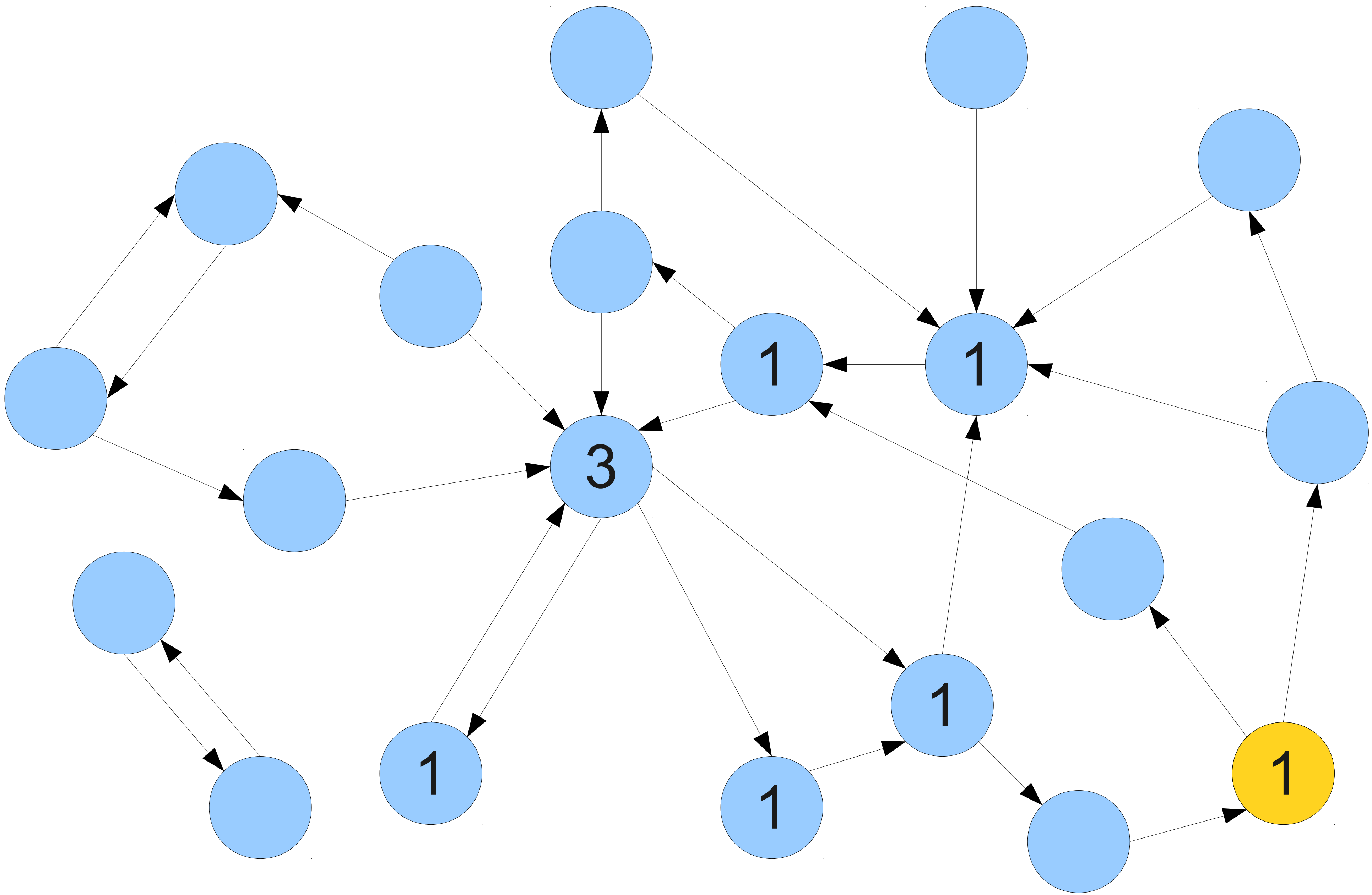
The Random Surfer Model



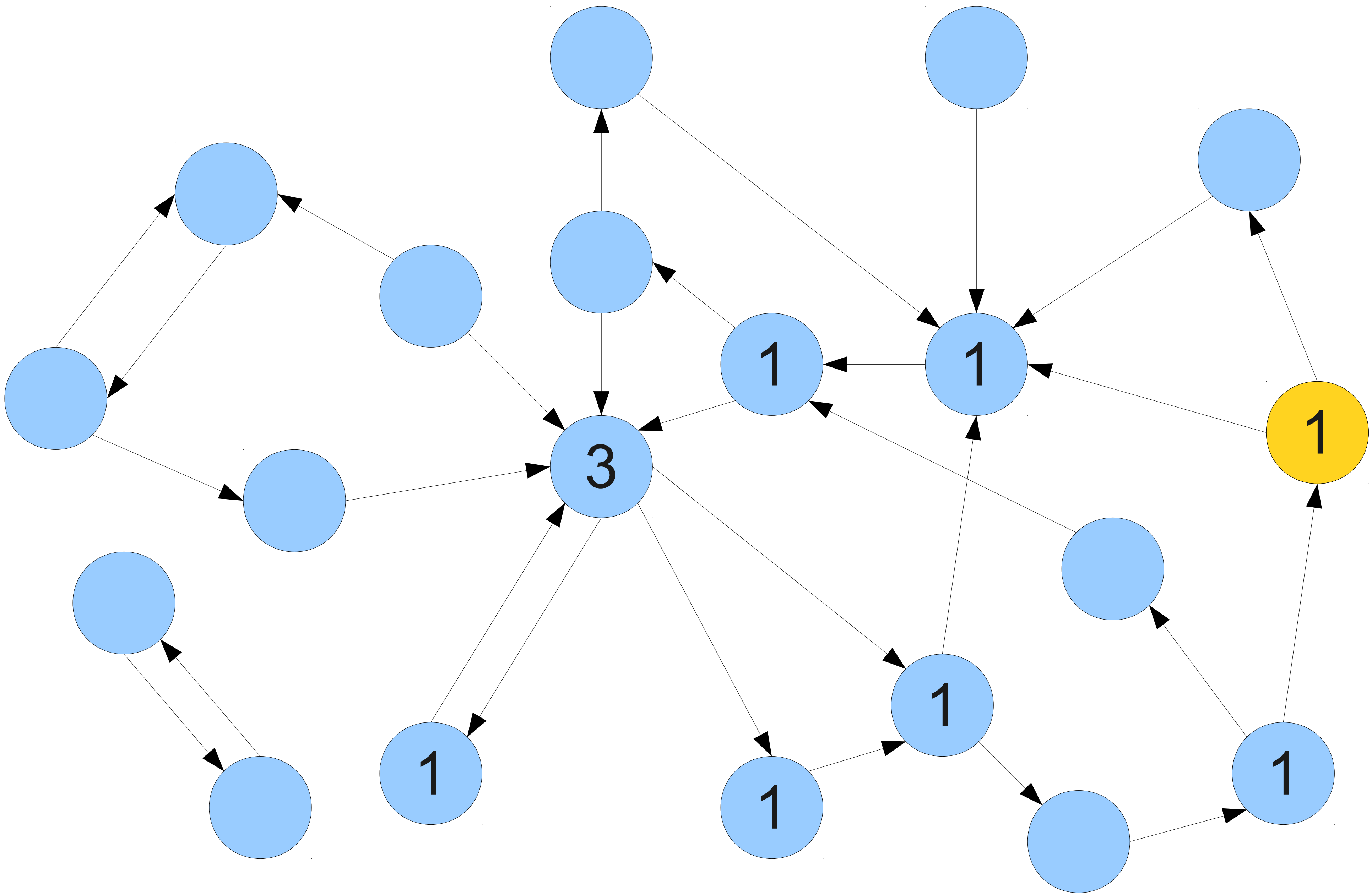
The Random Surfer Model



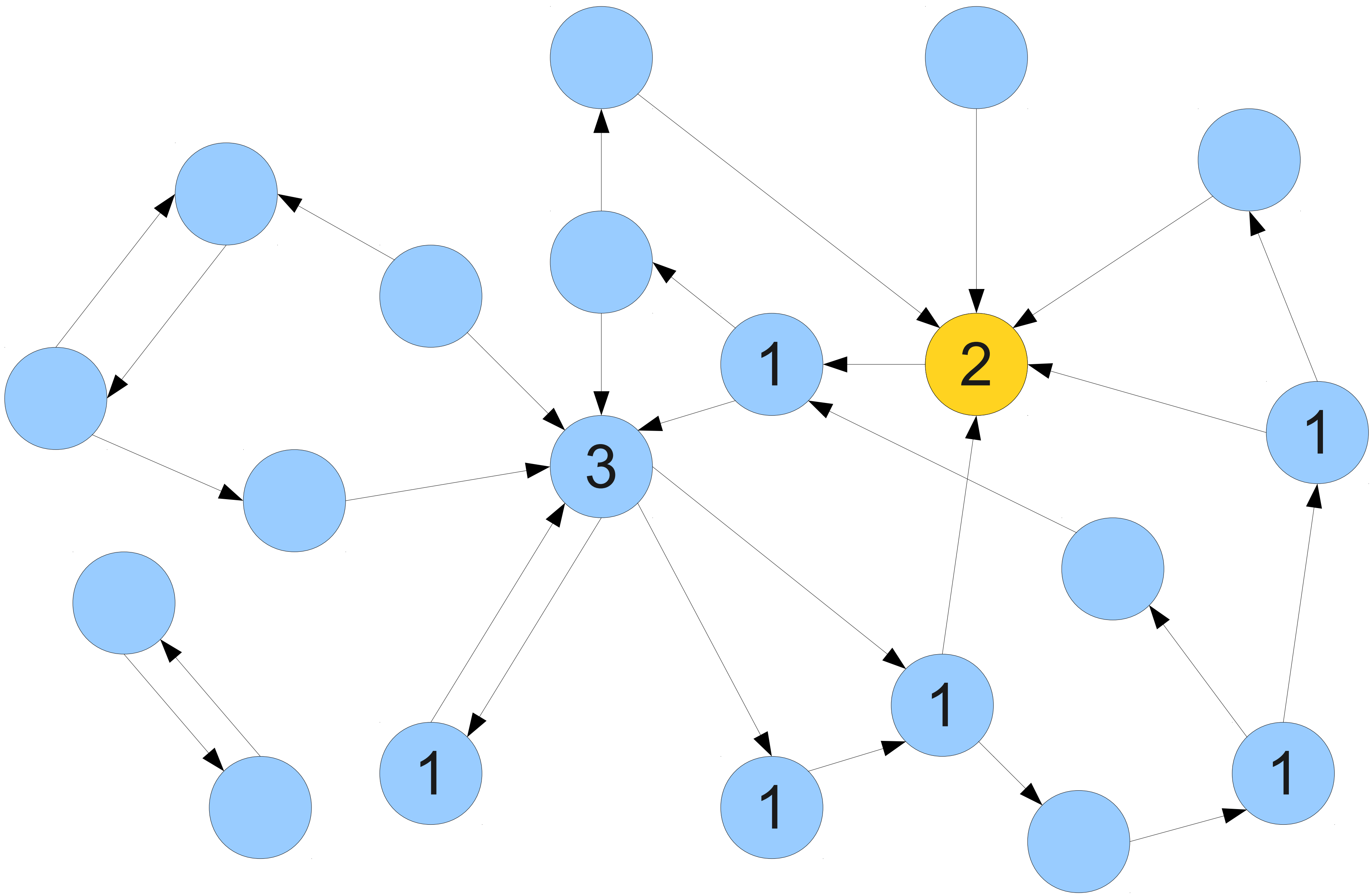
The Random Surfer Model



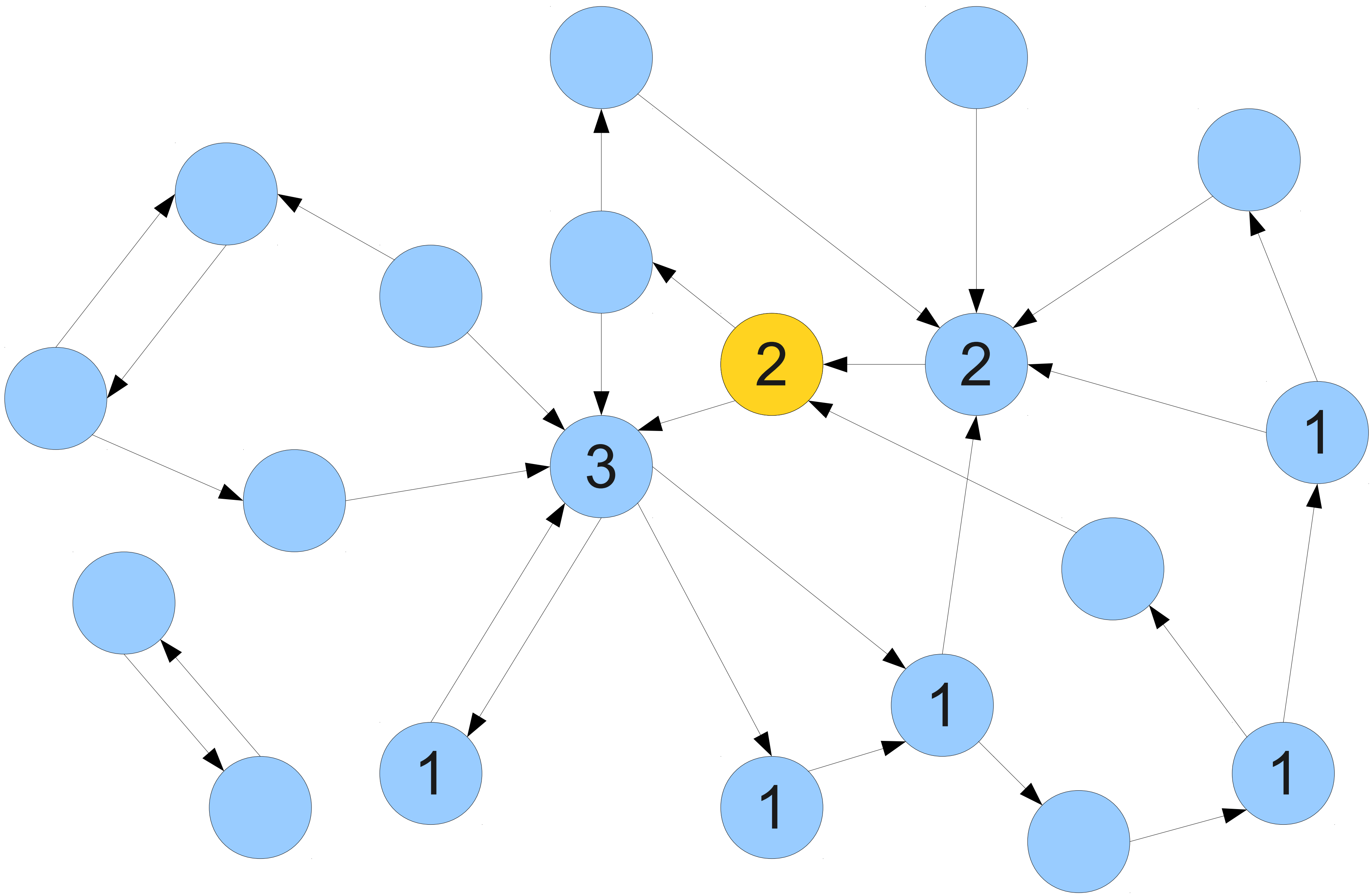
The Random Surfer Model



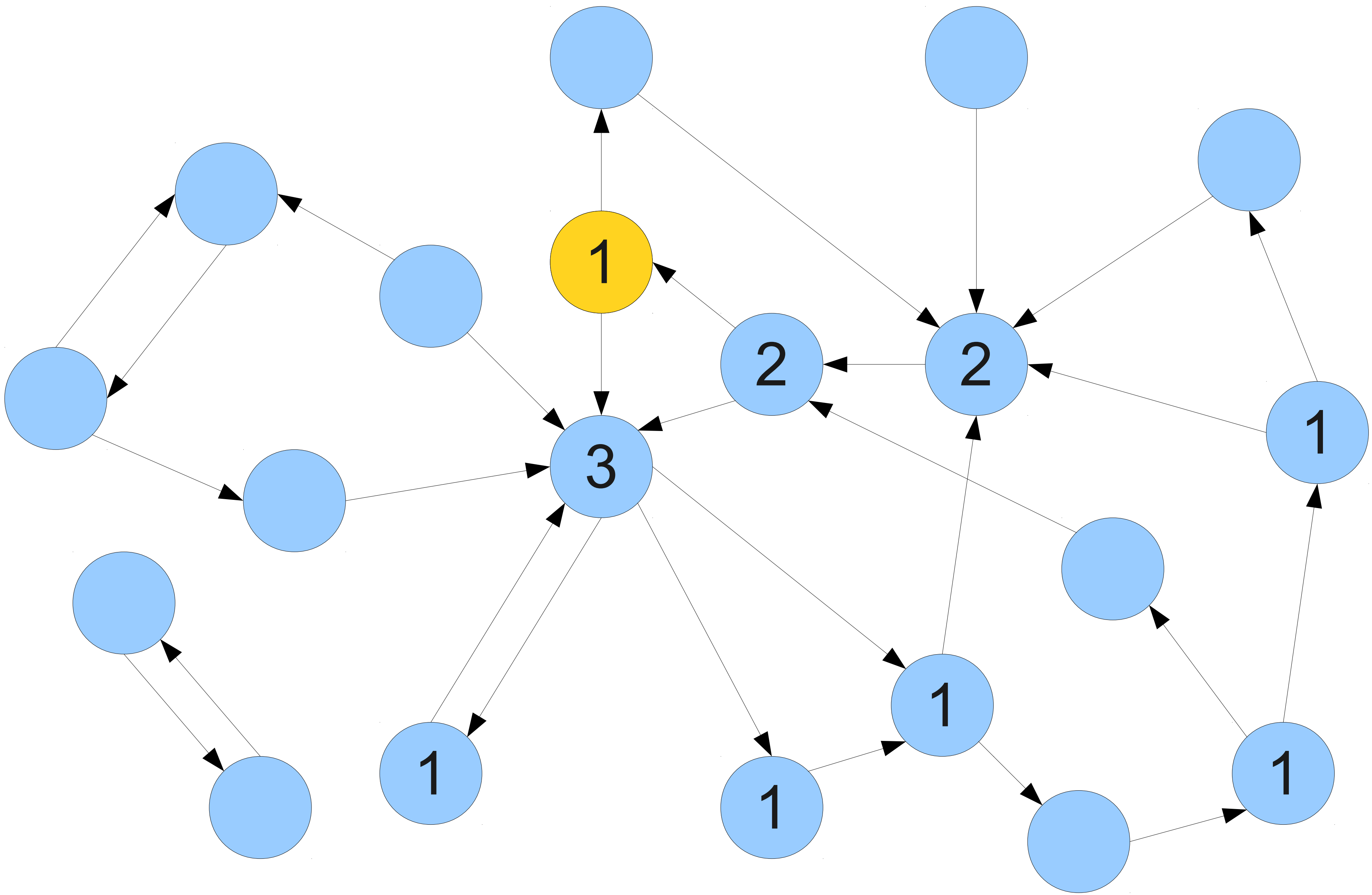
The Random Surfer Model



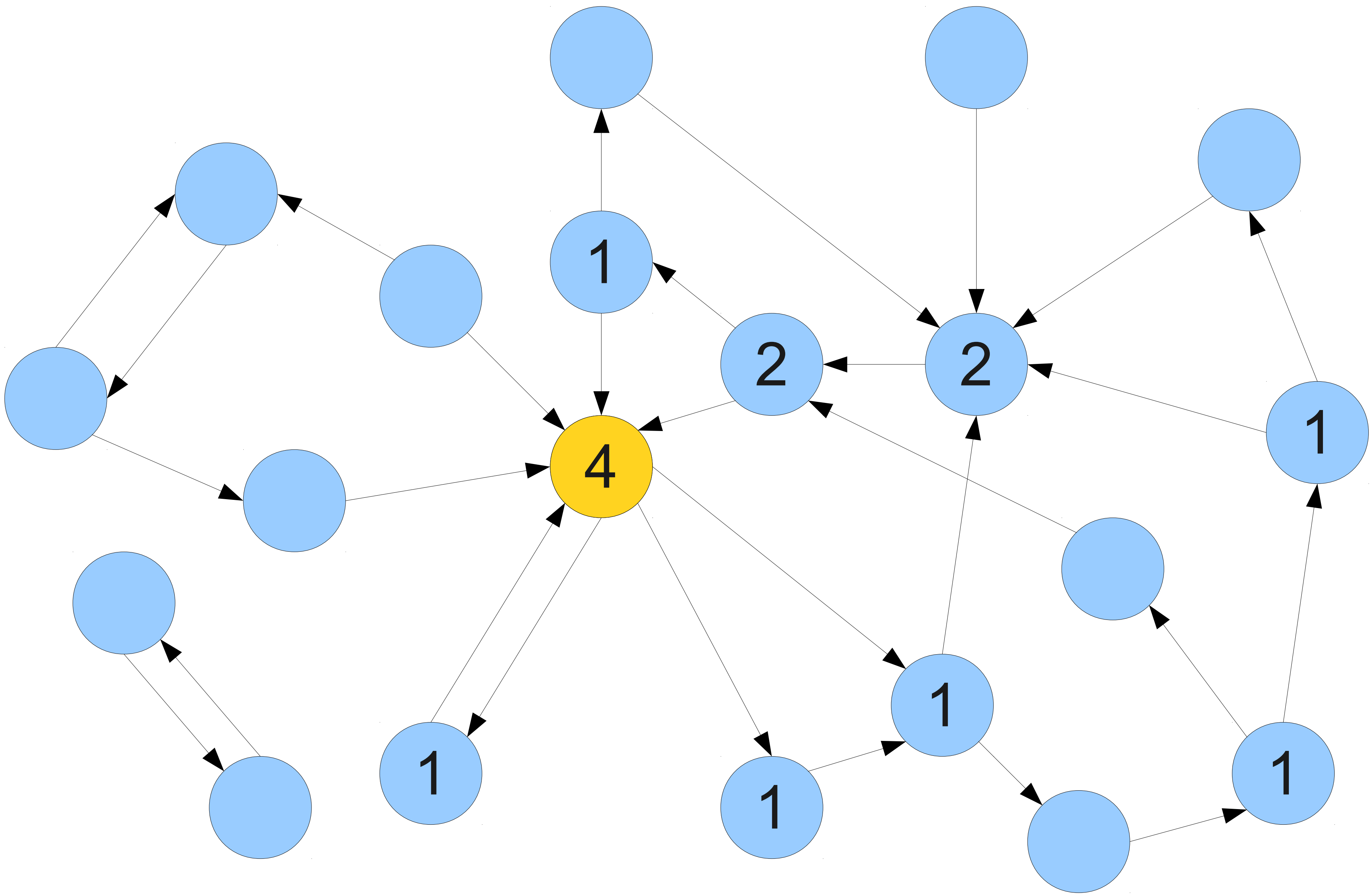
The Random Surfer Model



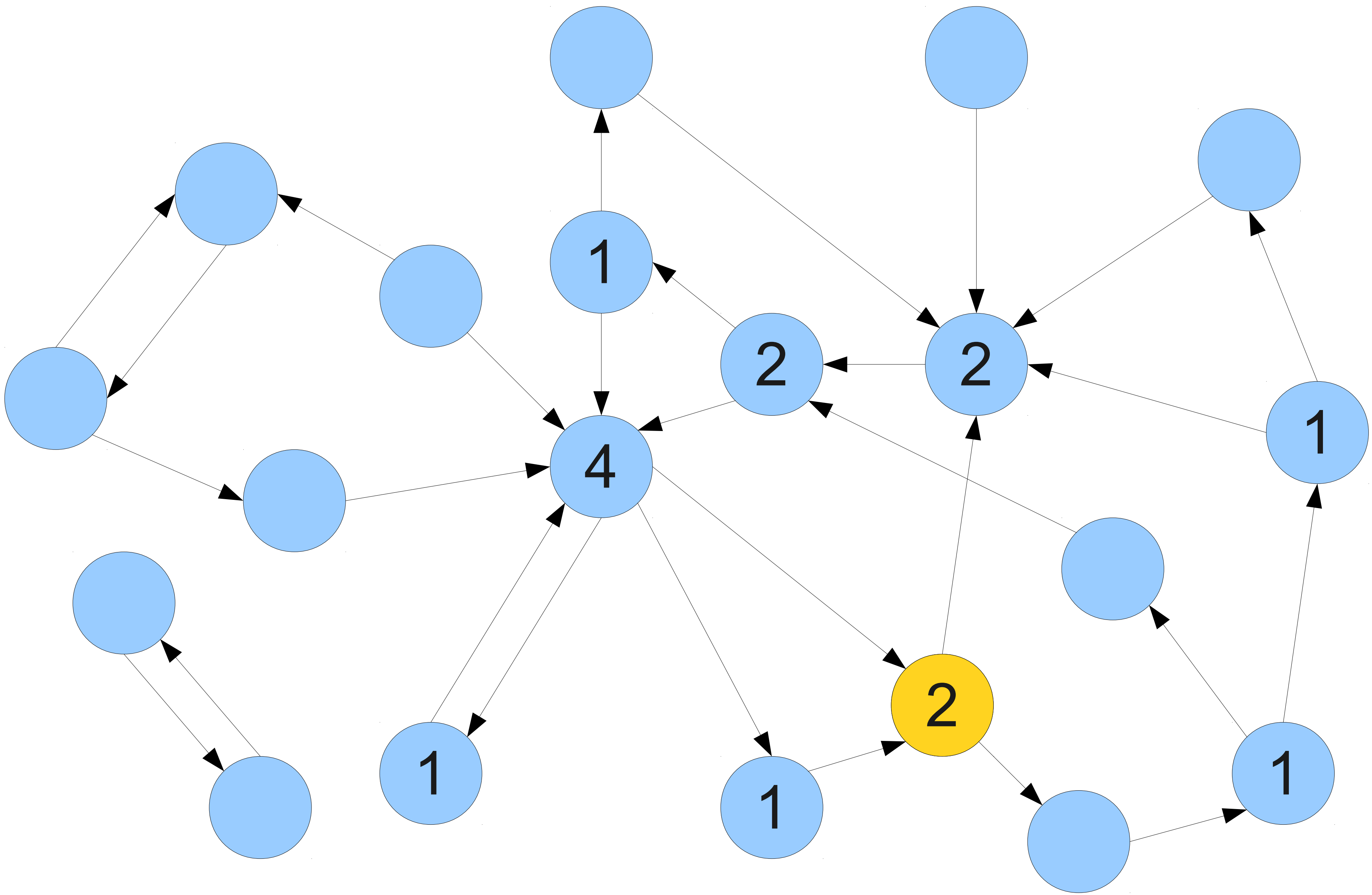
The Random Surfer Model



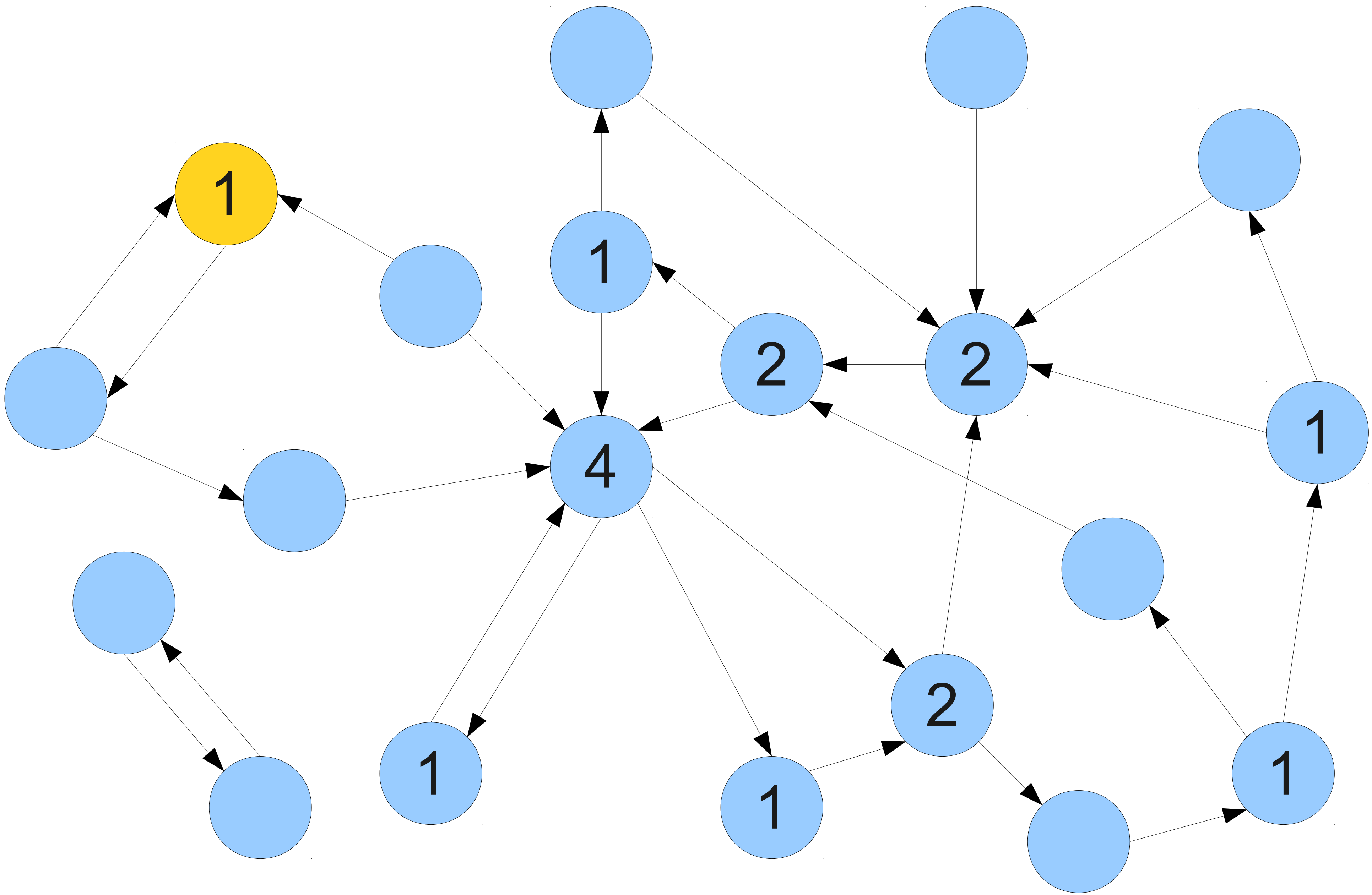
The Random Surfer Model



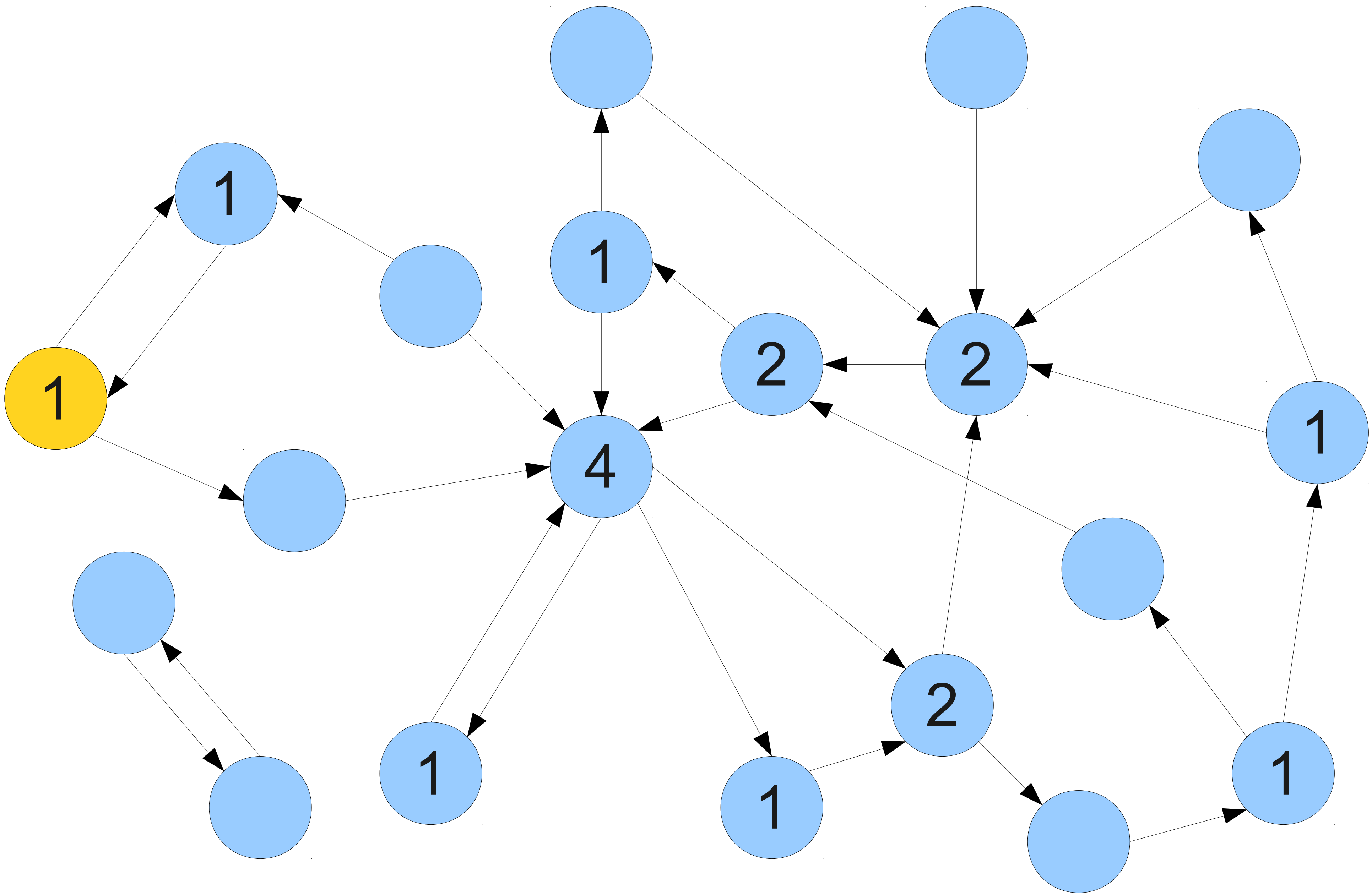
The Random Surfer Model



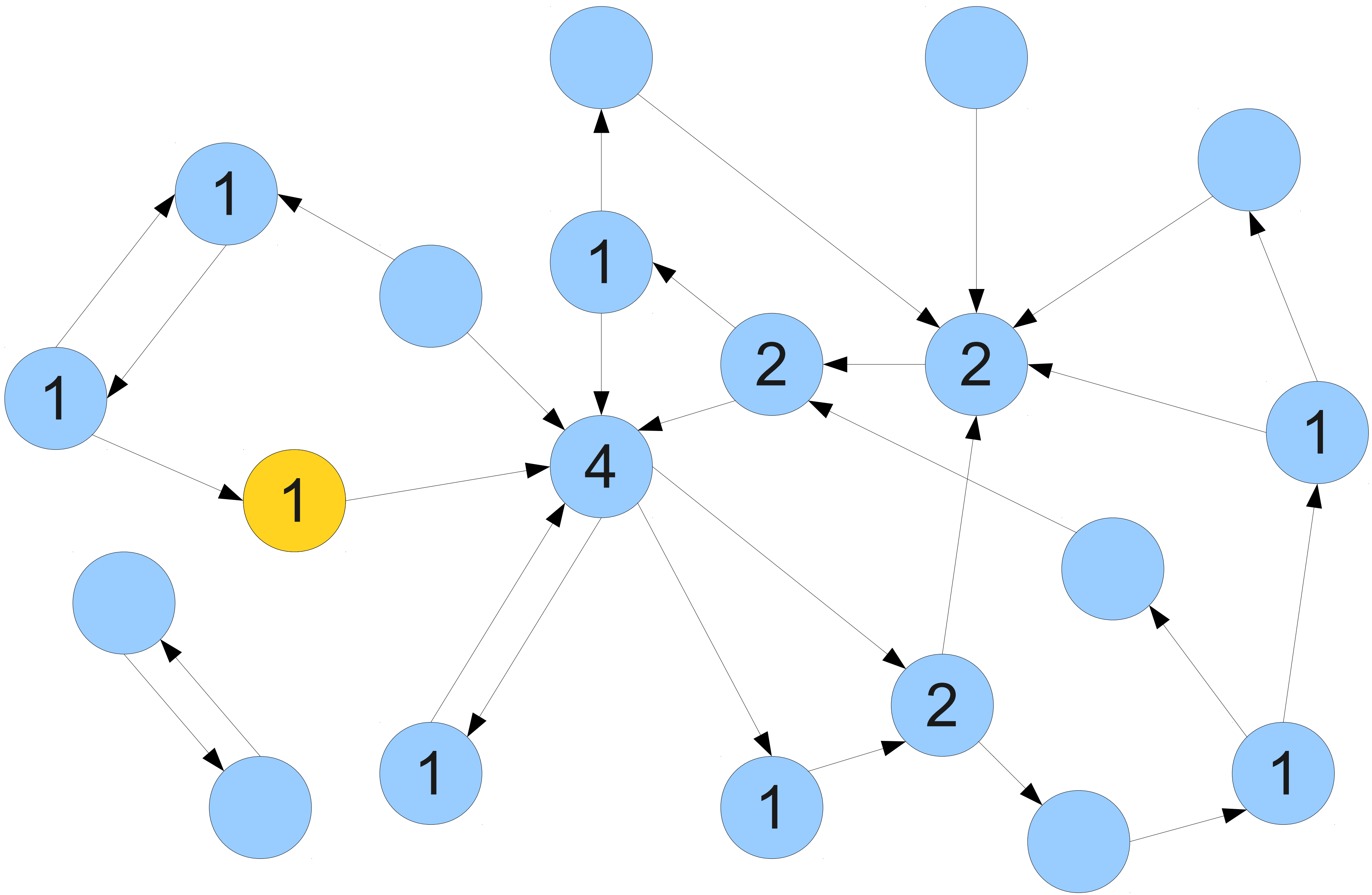
The Random Surfer Model



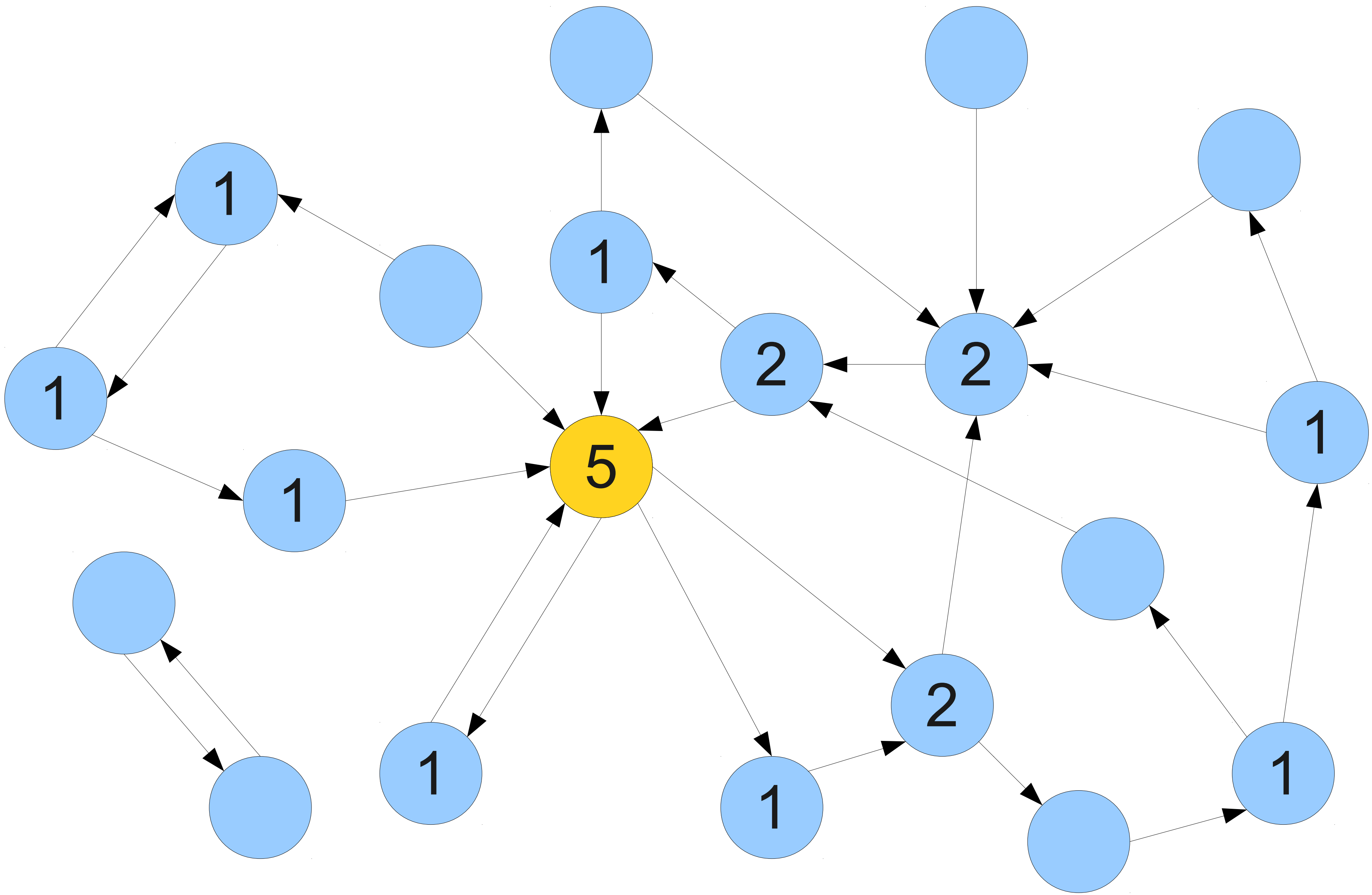
The Random Surfer Model



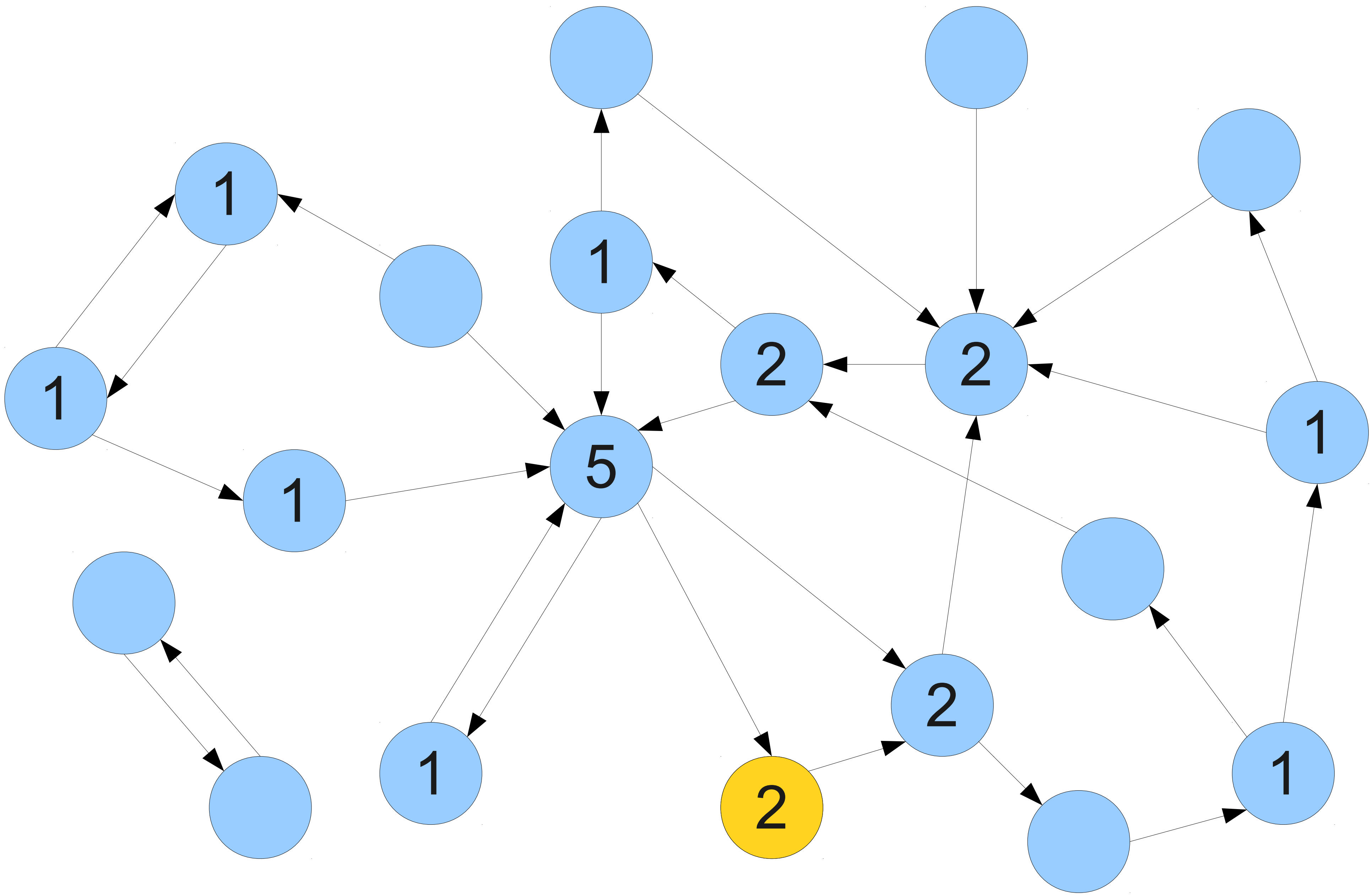
The Random Surfer Model



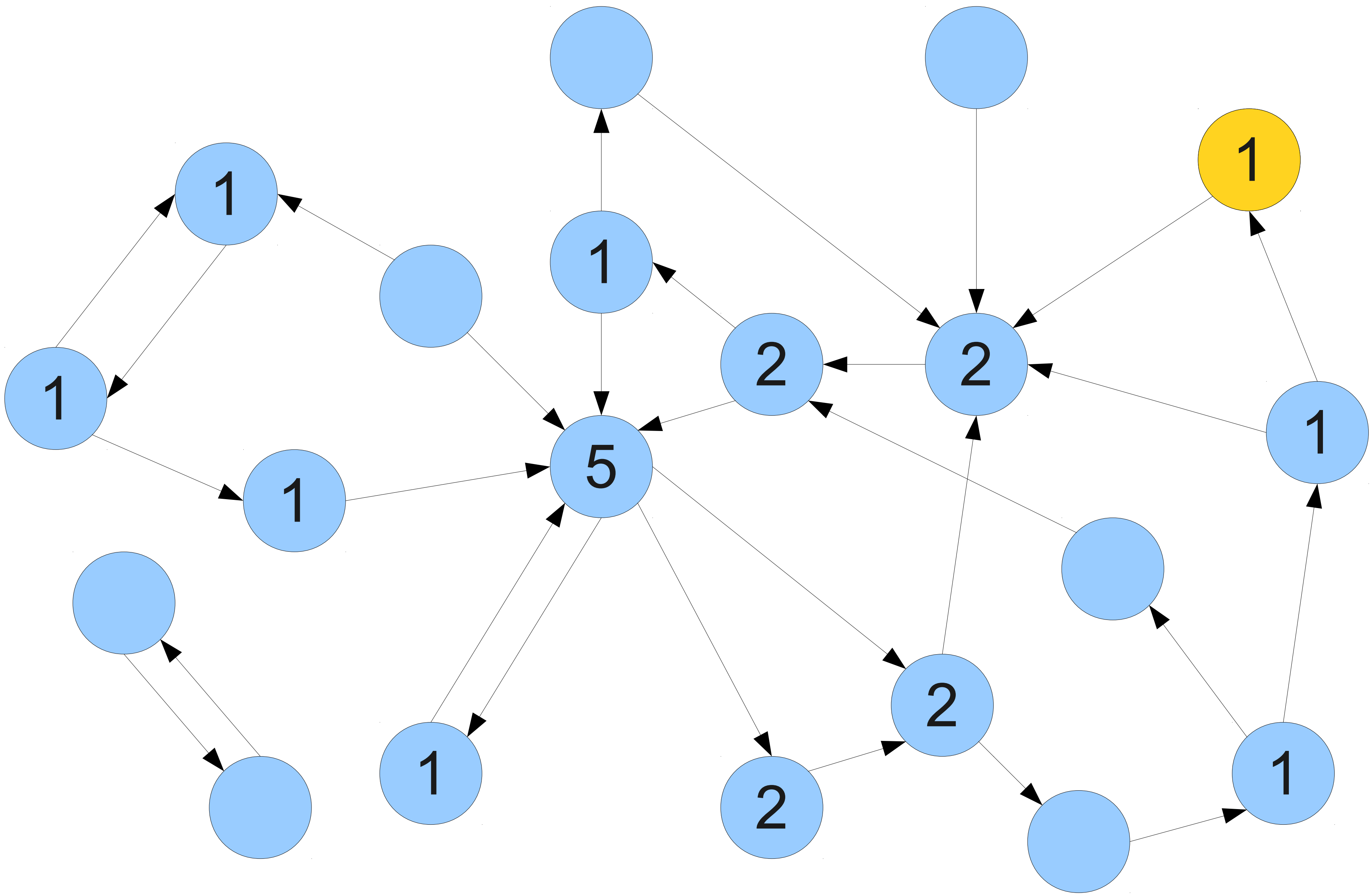
The Random Surfer Model



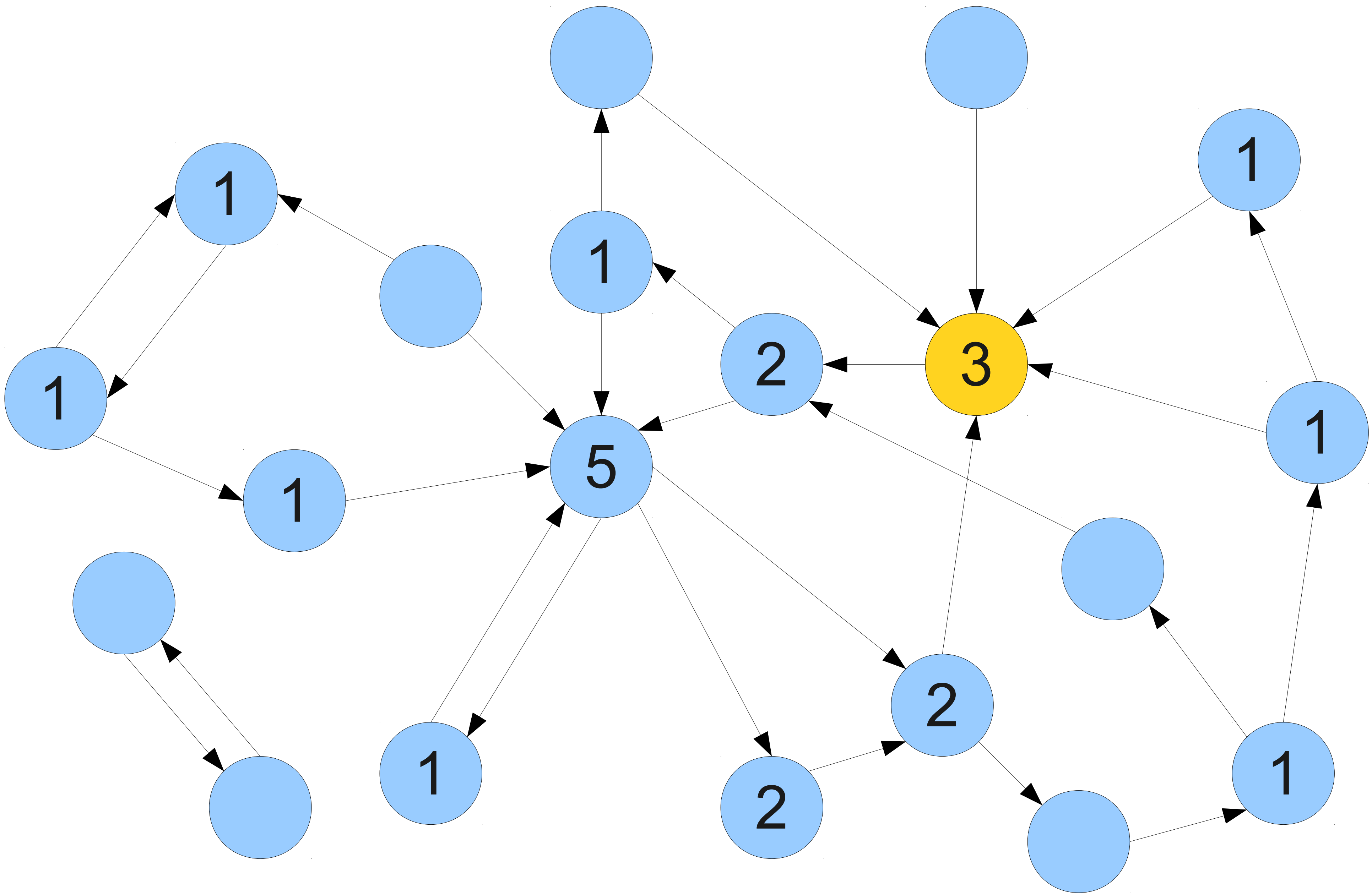
The Random Surfer Model



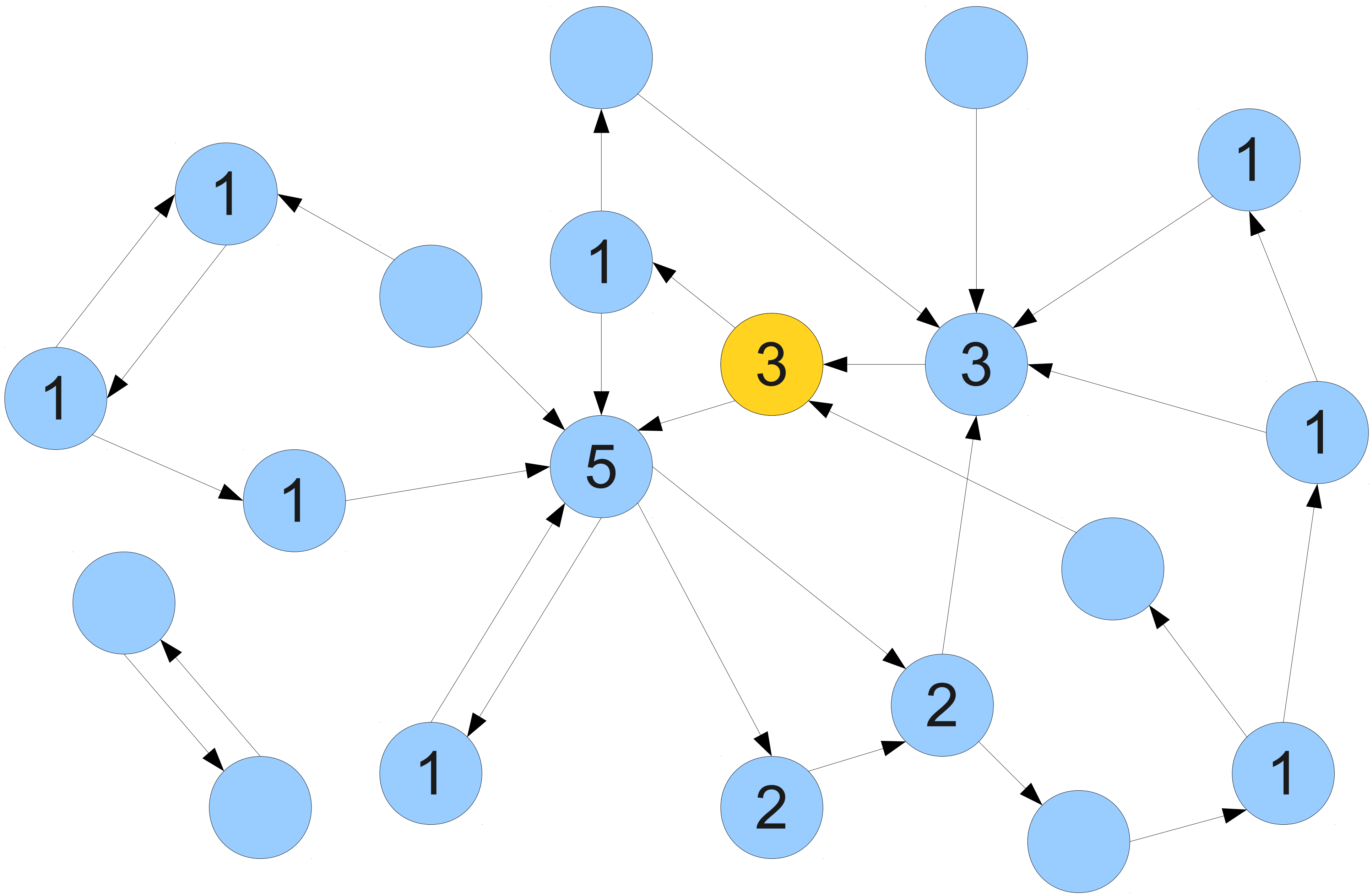
The Random Surfer Model



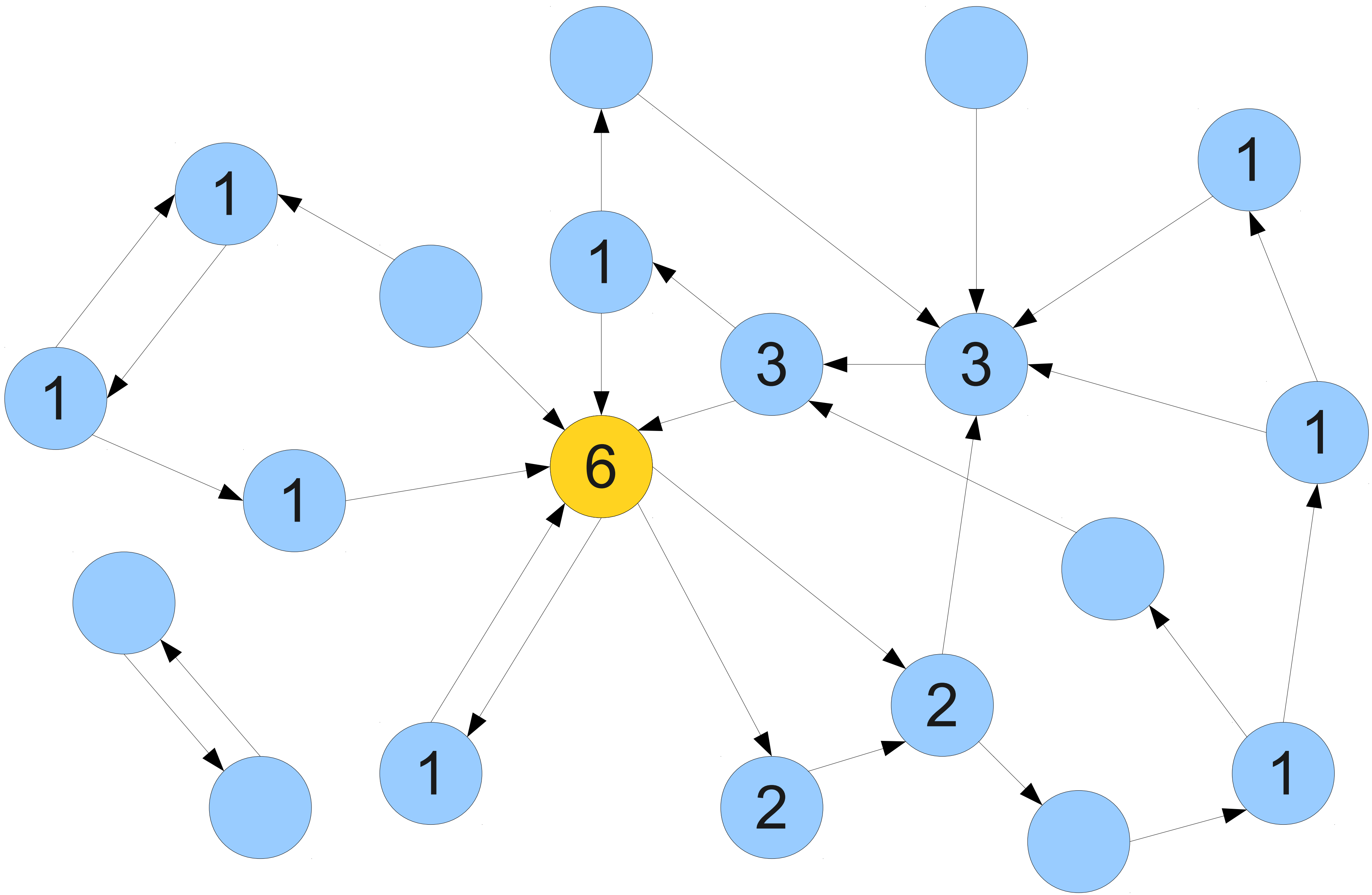
The Random Surfer Model



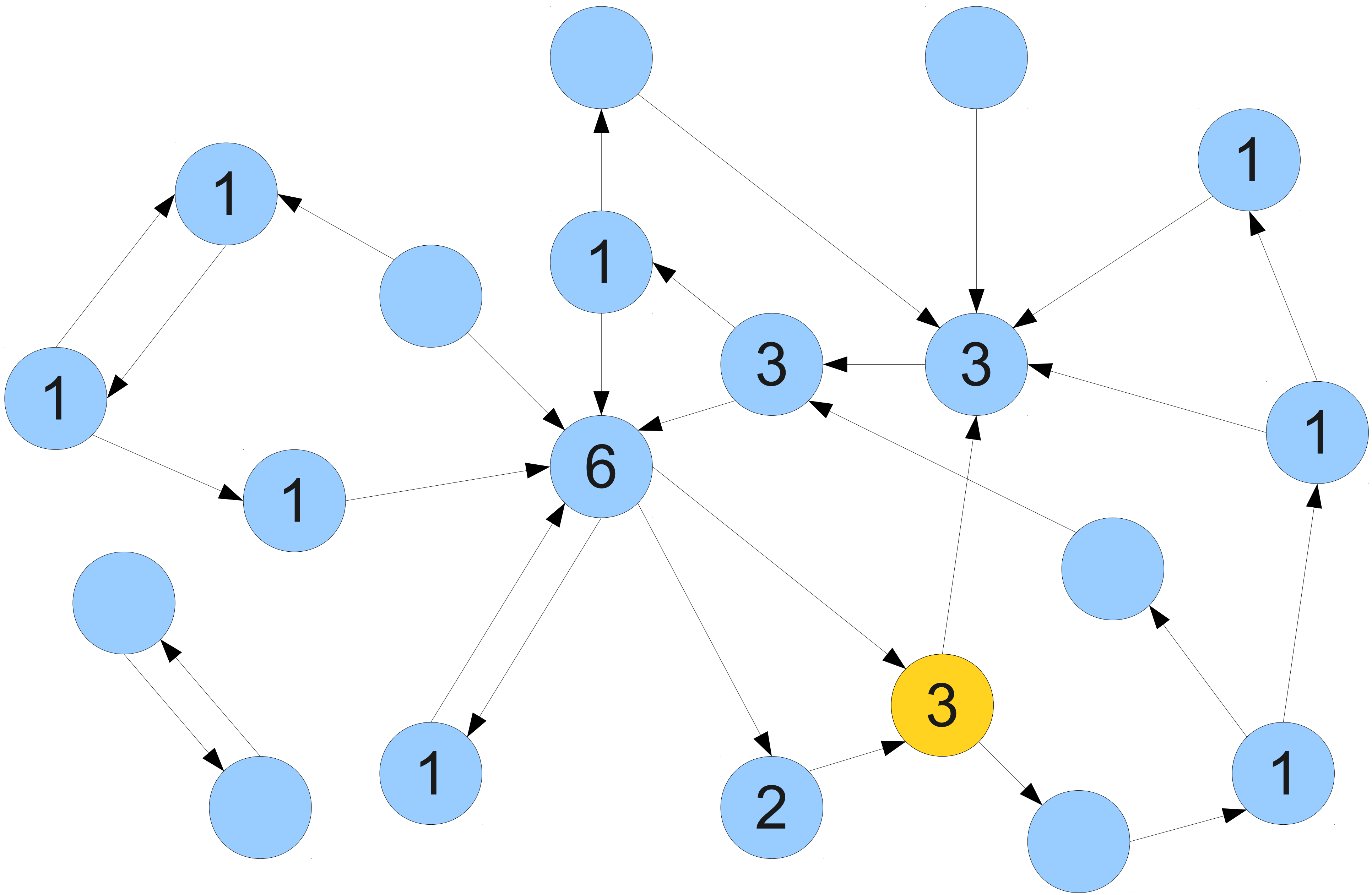
The Random Surfer Model



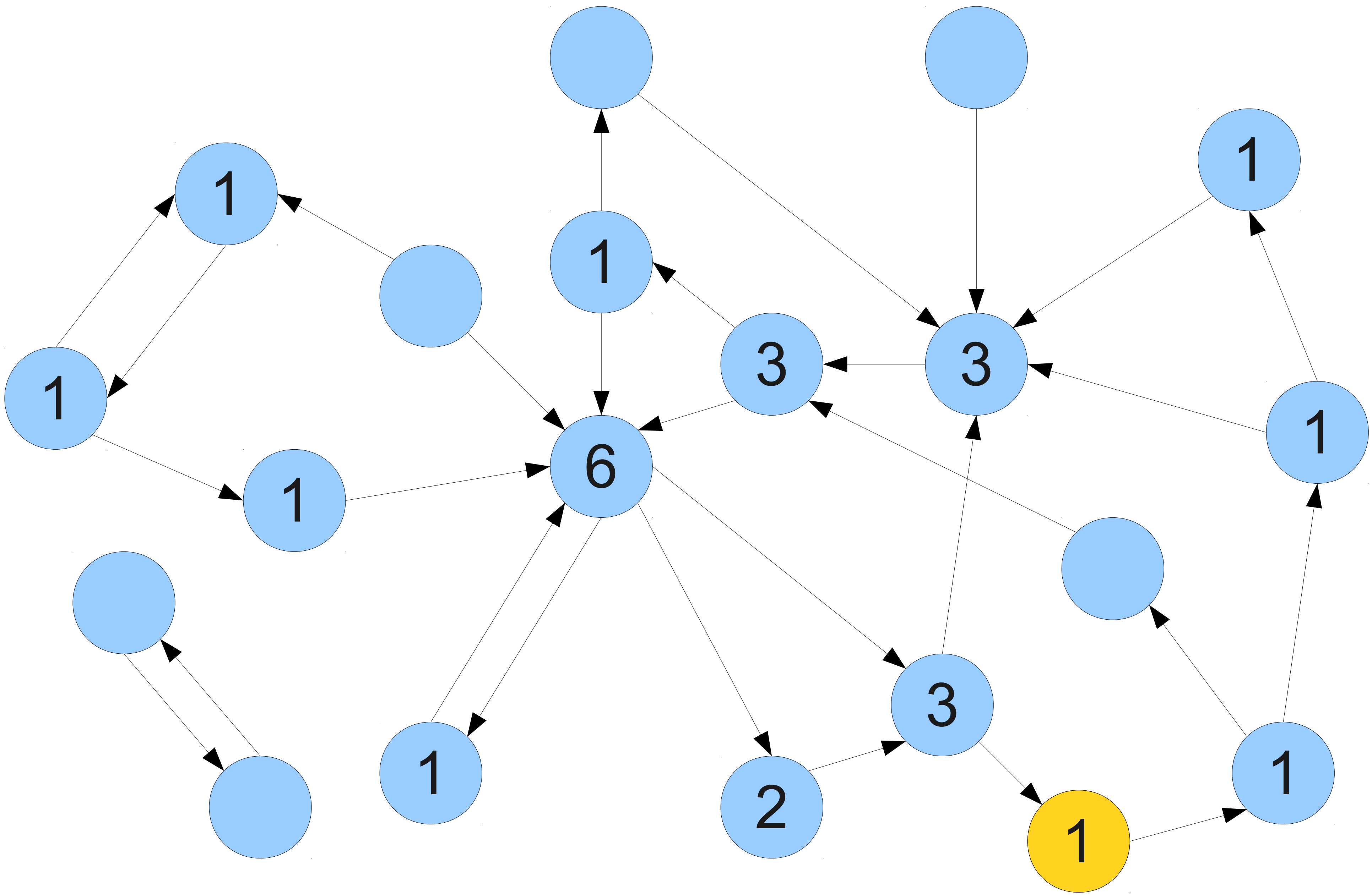
The Random Surfer Model



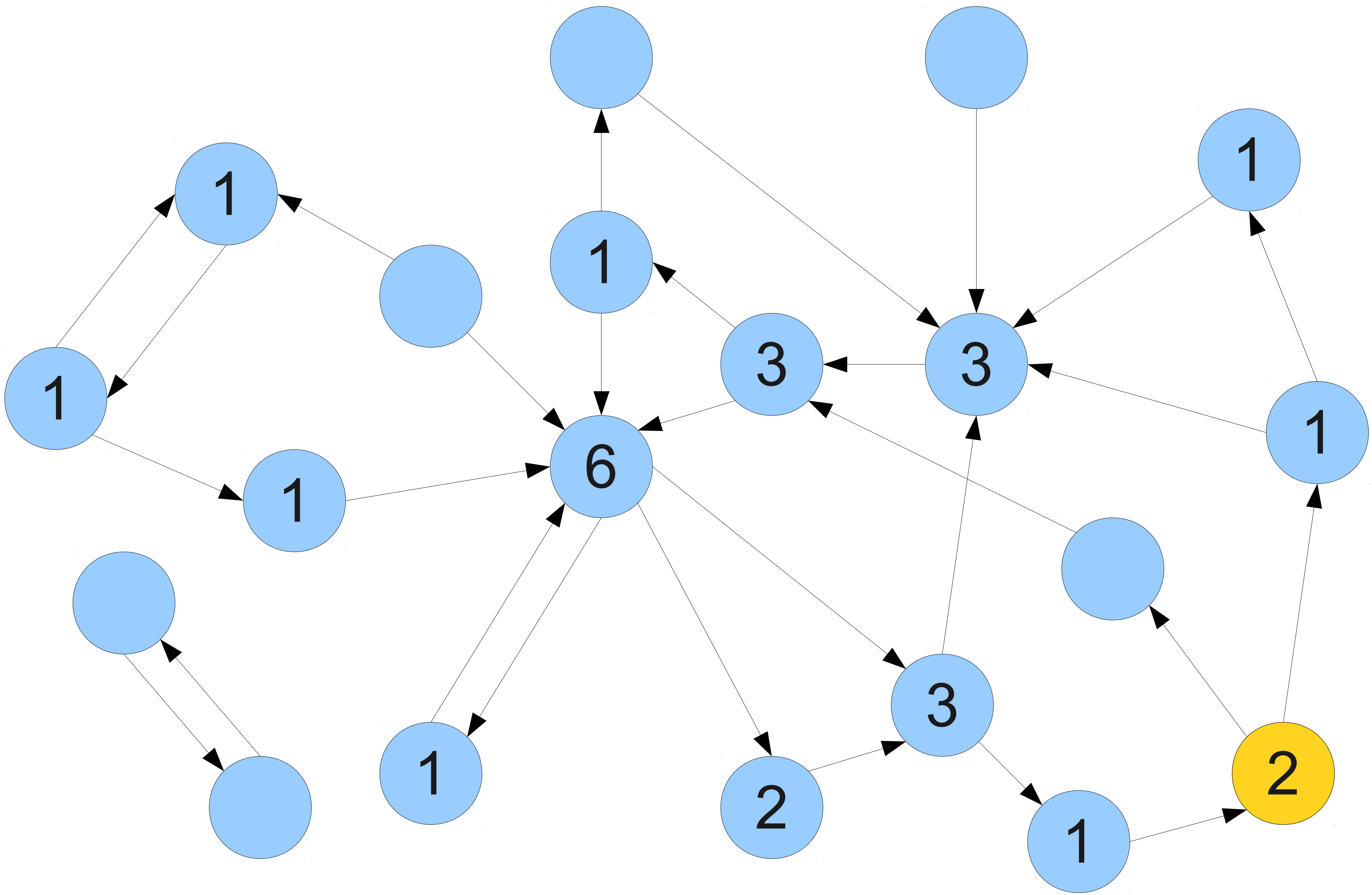
The Random Surfer Model



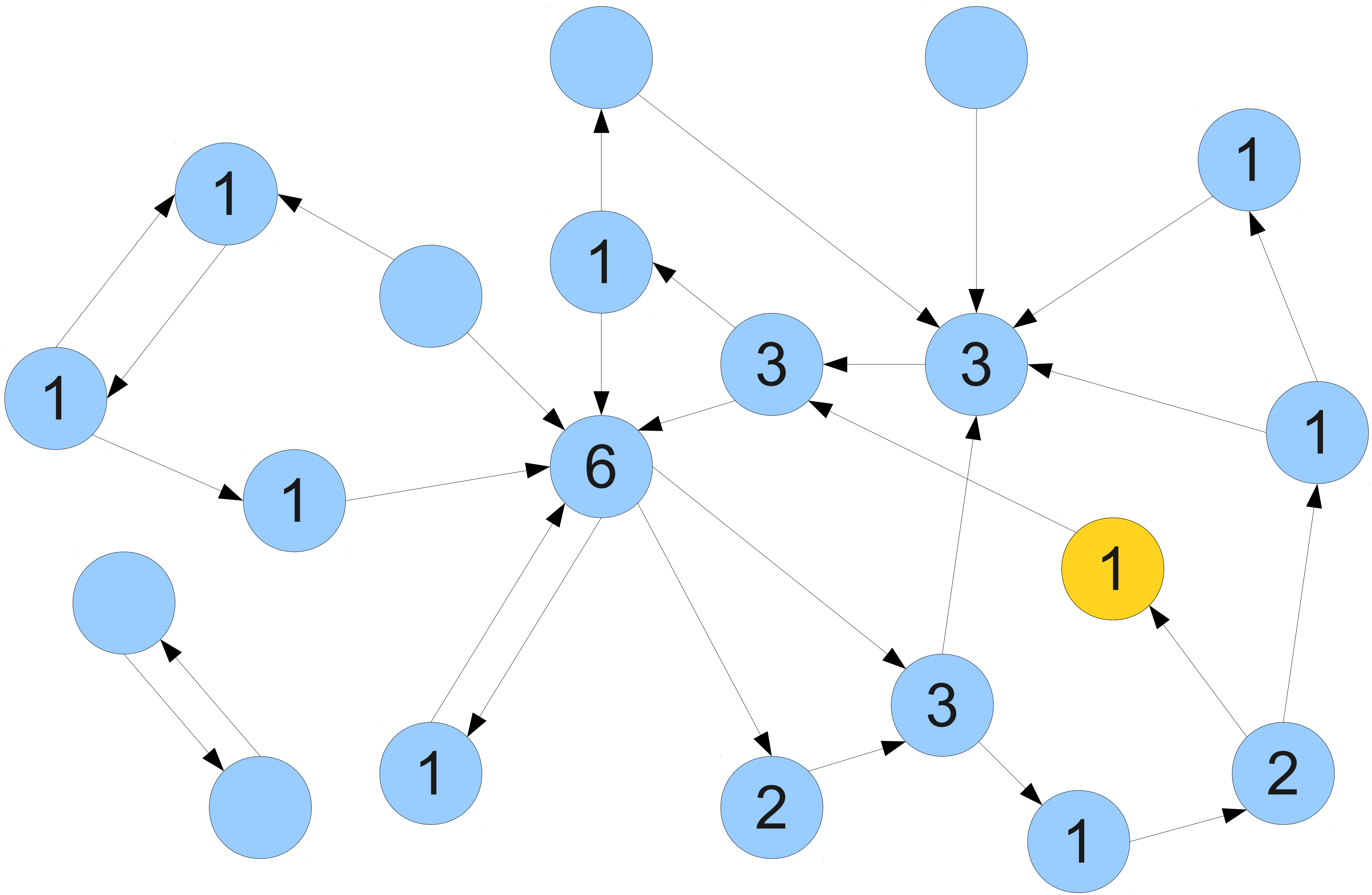
The Random Surfer Model



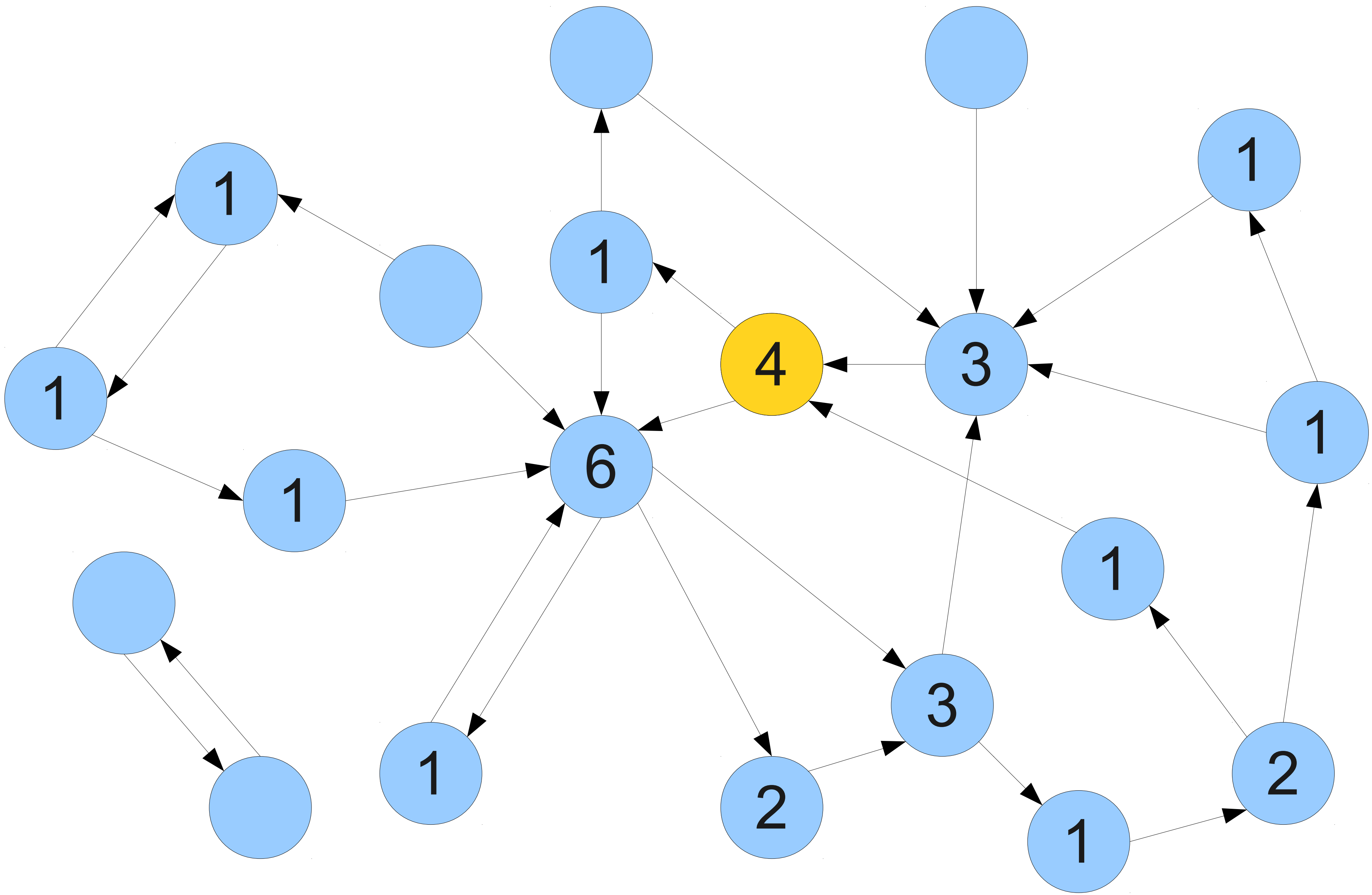
The Random Surfer Model



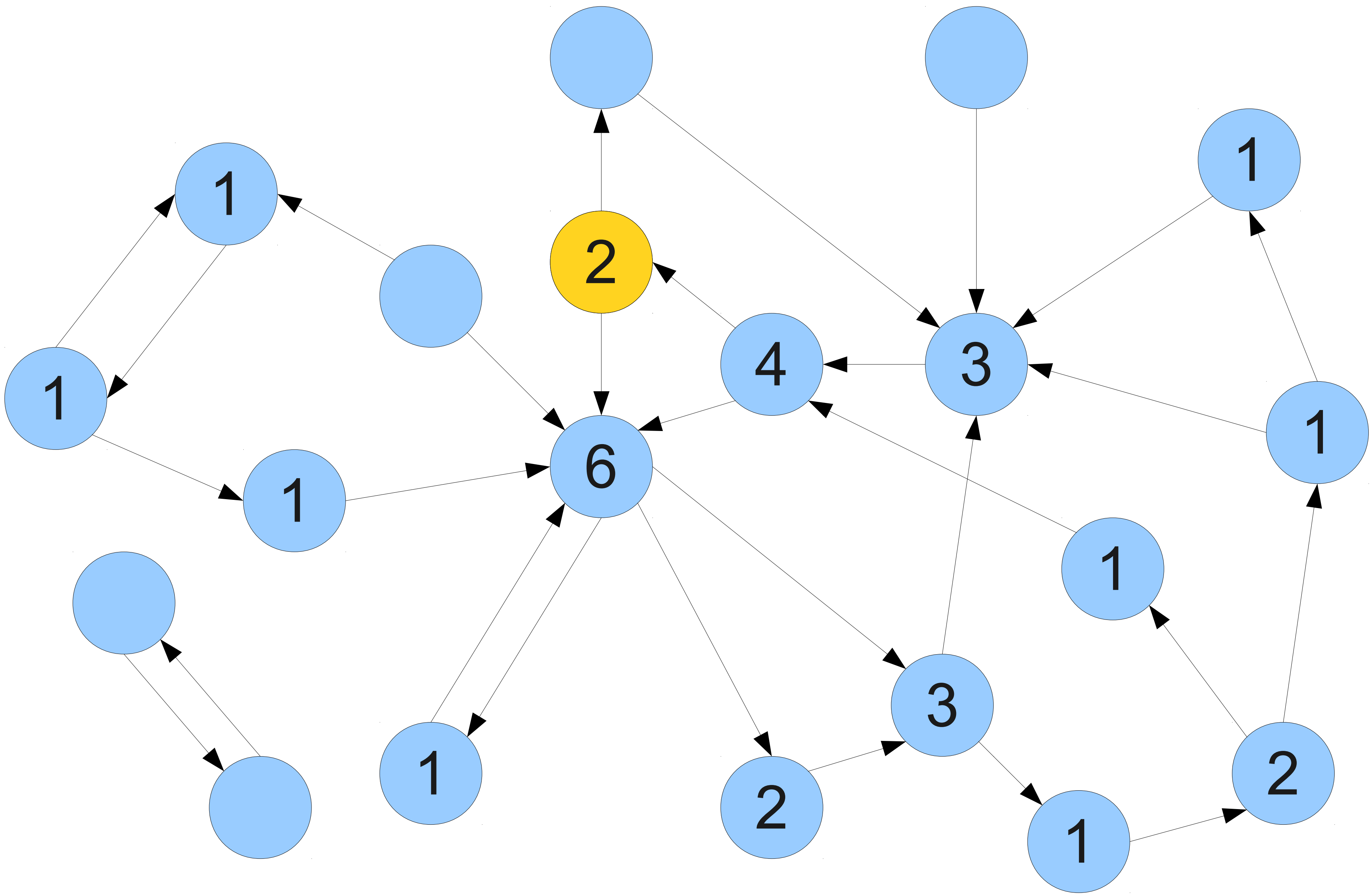
The Random Surfer Model



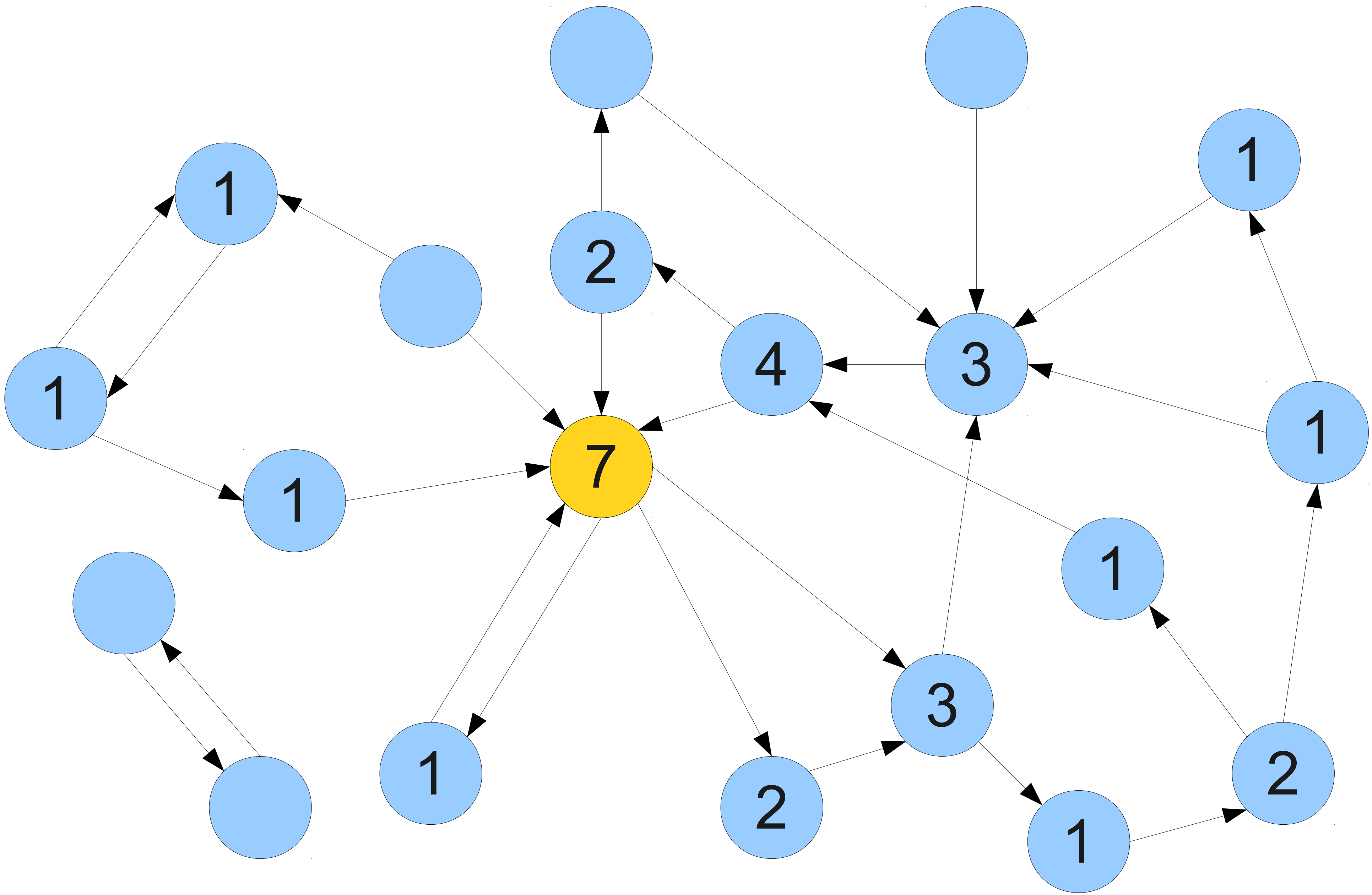
The Random Surfer Model



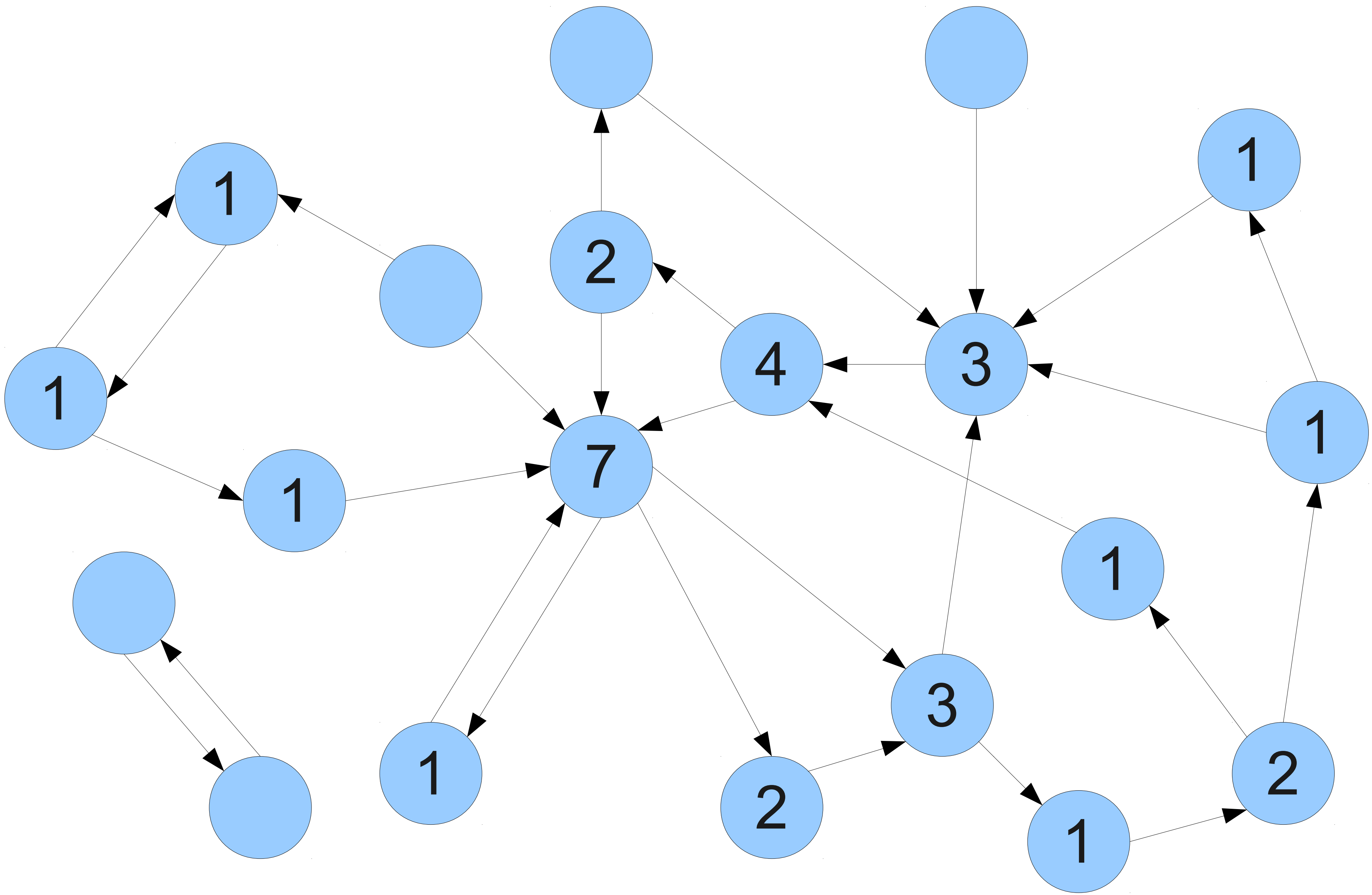
The Random Surfer Model



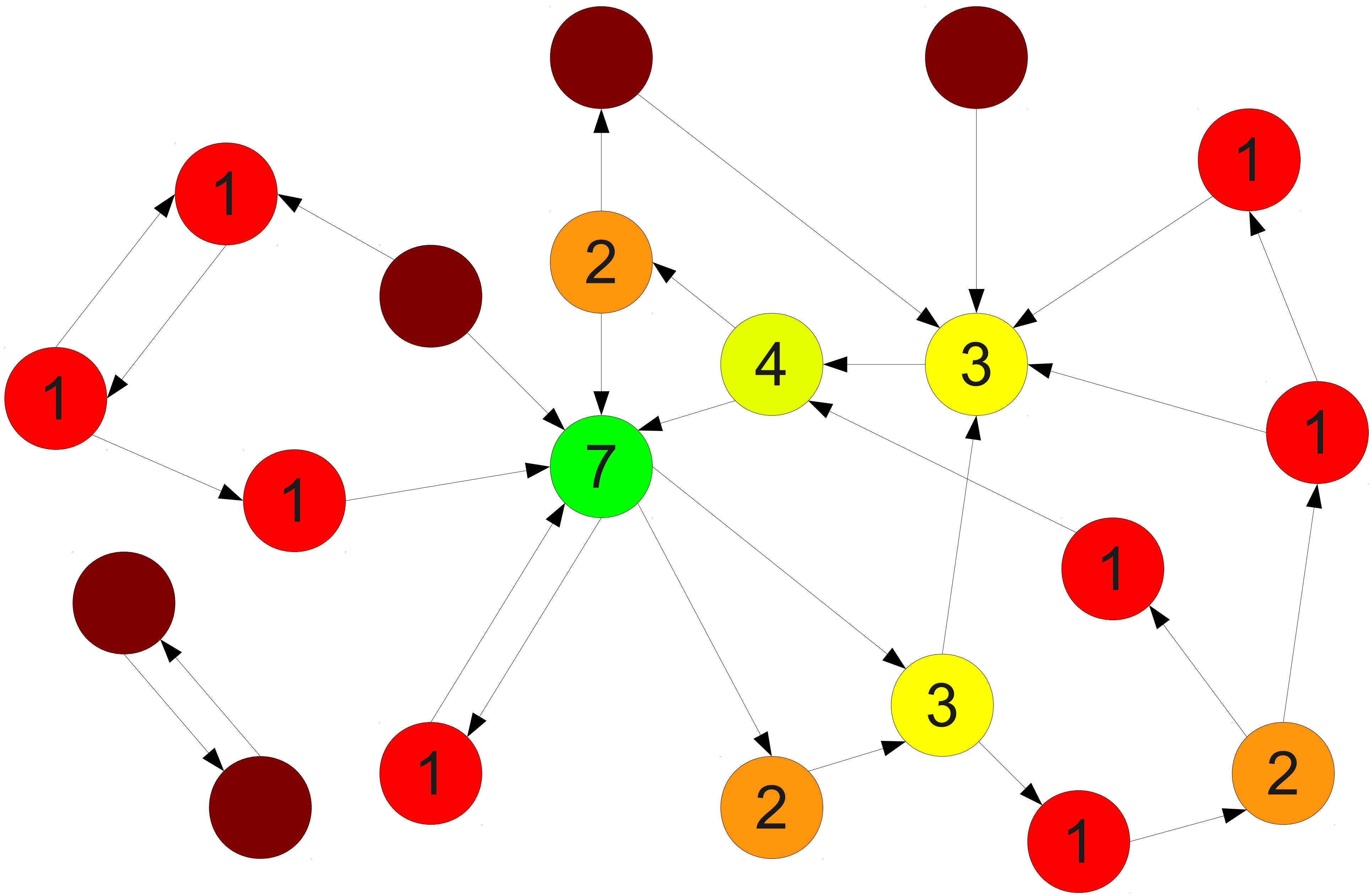
The Random Surfer Model



The Random Surfer Model



The Random Surfer Model



Ranking Articles with the RSM

- Randomly walk through the graph.
- At each step, either
  - Jump to a totally random article, or
  - Follow a random link.
- Record how many times each article was visited.
- The most-visited articles are, in some sense, the most important.

Other Applications of the RSM

- Ecosystem Stability:
  - Each node represents a species.
  - Edges represent one species that eats another.
  - High-value species are those that are important to the stability of the ecosystem.
- Learn more:
  - **http://news.bbc.co.uk/2/hi/8238462.stm**

Who invented this?

[Our approach] can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page.

[Our approach] can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its **PageRank**.

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page
Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

http://quotingquotes.co.uk/wp-content/uploads/2012/01/1-larry-page_sergey-brin.jpg

Great things are possible in computing.

Great things are possible in computing.

You just need to do a little random surfing.