Category-based routing in social networks

Denitsa Dobreva

Overview

- Small-world phenomenon
- Definitions and routing algorithm
- Building categories
- Summary

Part 1

Small-world phenomenon

Small world phenomenon

- The phenomenon of elements being perceived to be very far away but are only at a distance of few hops

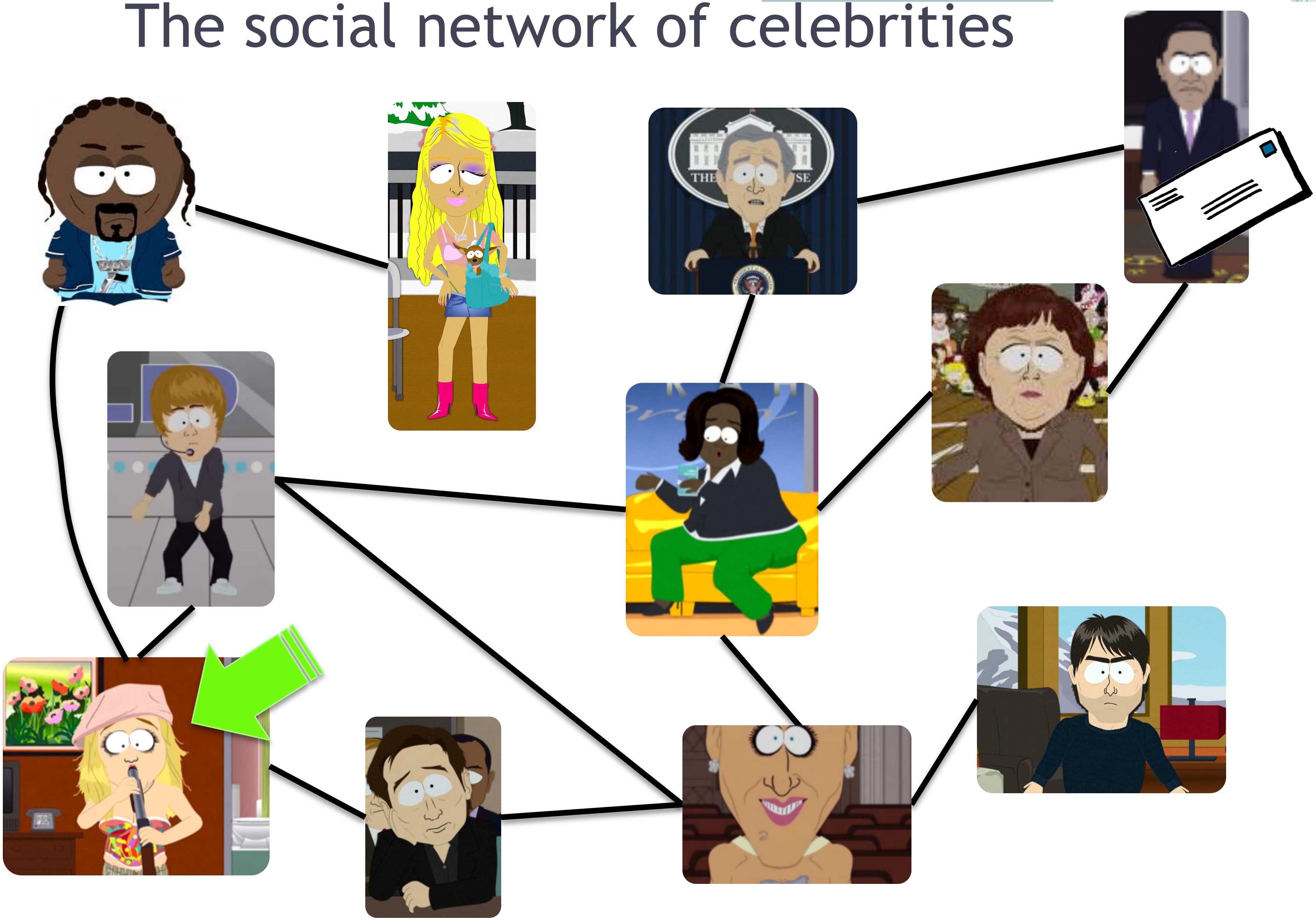
- The experience of meeting a complete stranger and finding out you share a mutual friend

  - "It's a small world" … we say

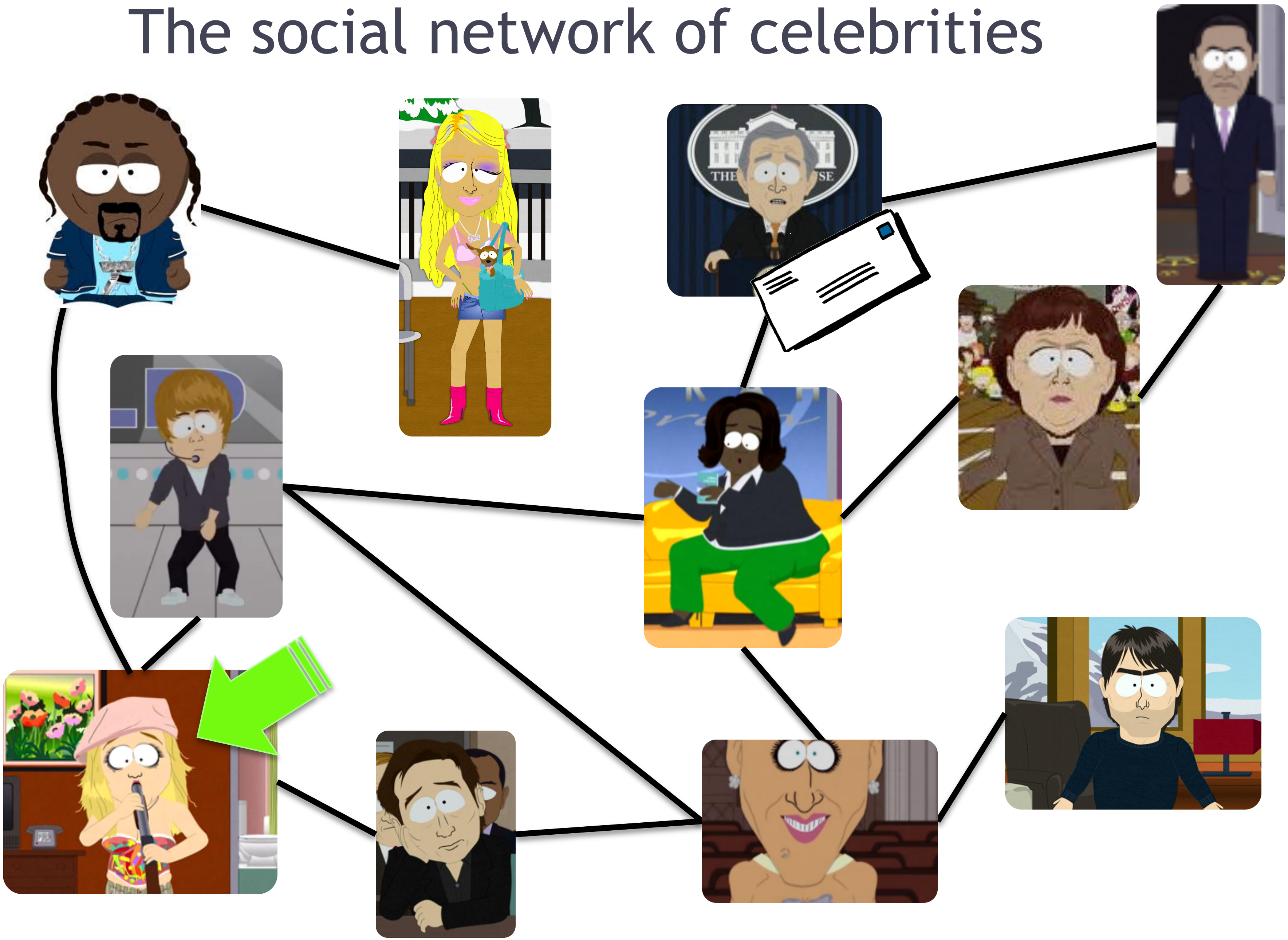
The notion of the small world phenomenon

- Milgram's experiment form the 1960's
  - Given: Social network
  - Idea: How many jumps are needed to reach a random person?
  - Starters: Random people from Nebraska
  - Target: Person, who lived in Massachusetts and worked in Boston
  - Known: Basic information about the target
  - Rule: Send to a people known on first name basis
- Routing using only local information and simple facts

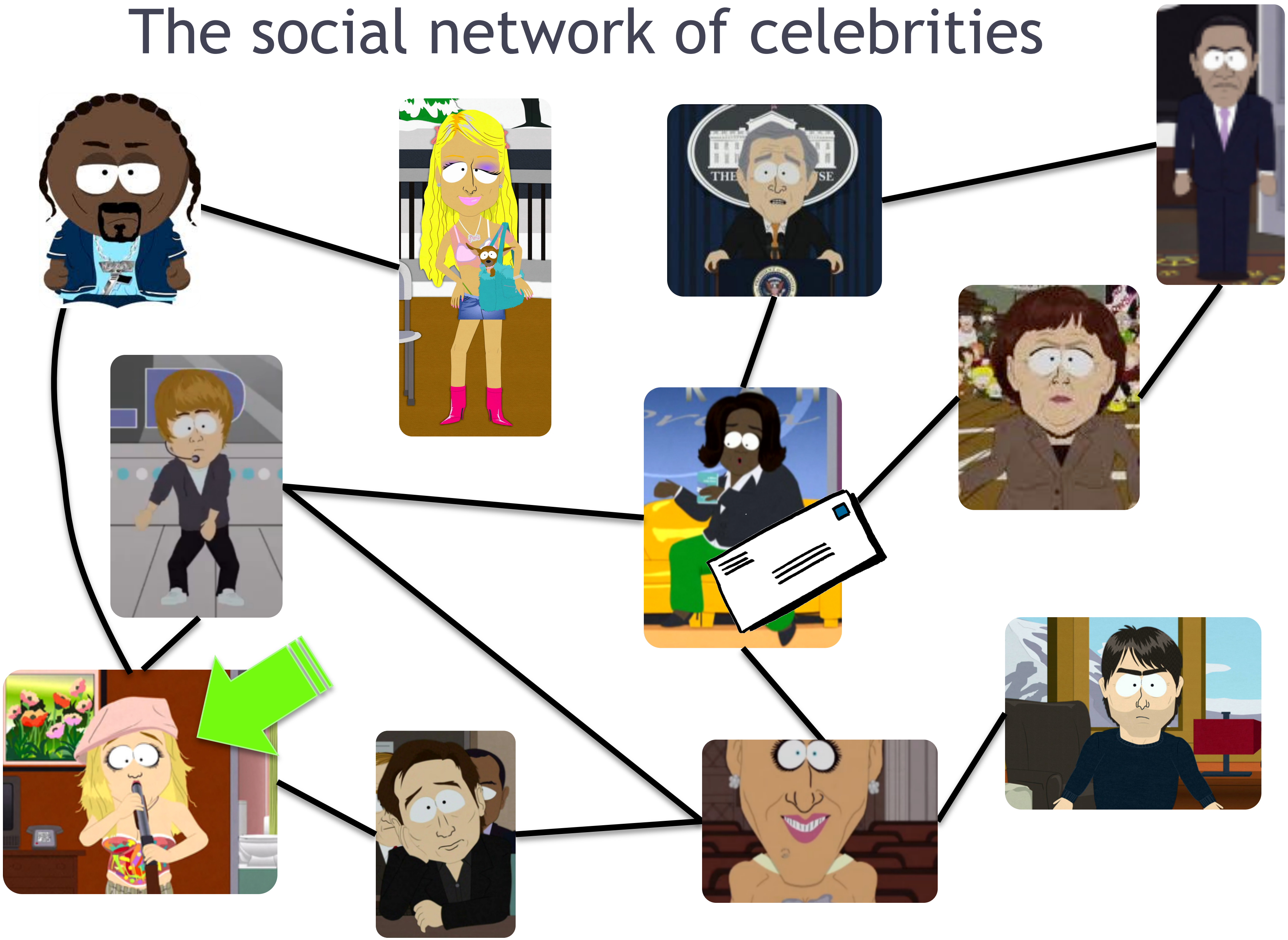
The social network of celebrities



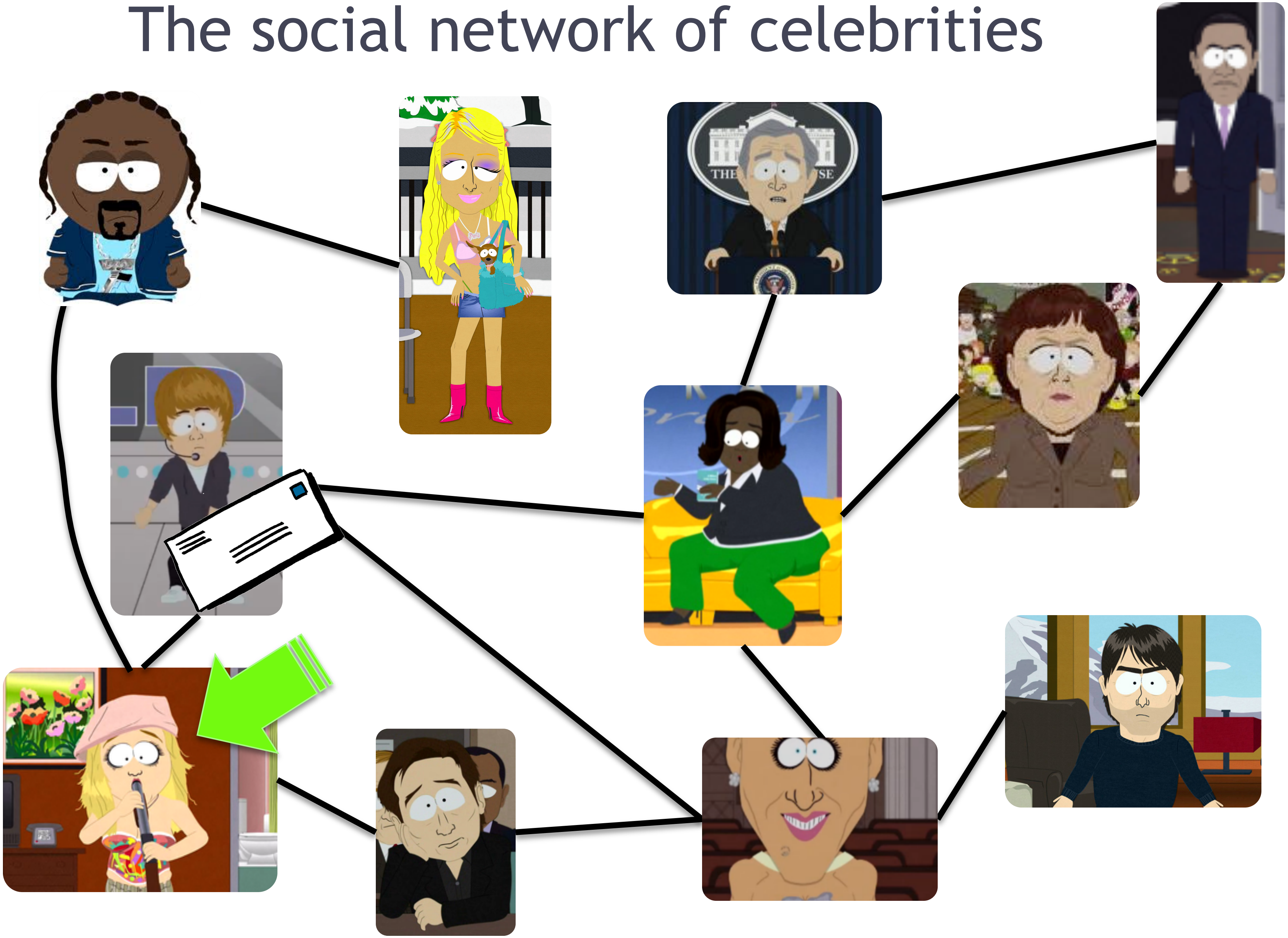
The social network of celebrities



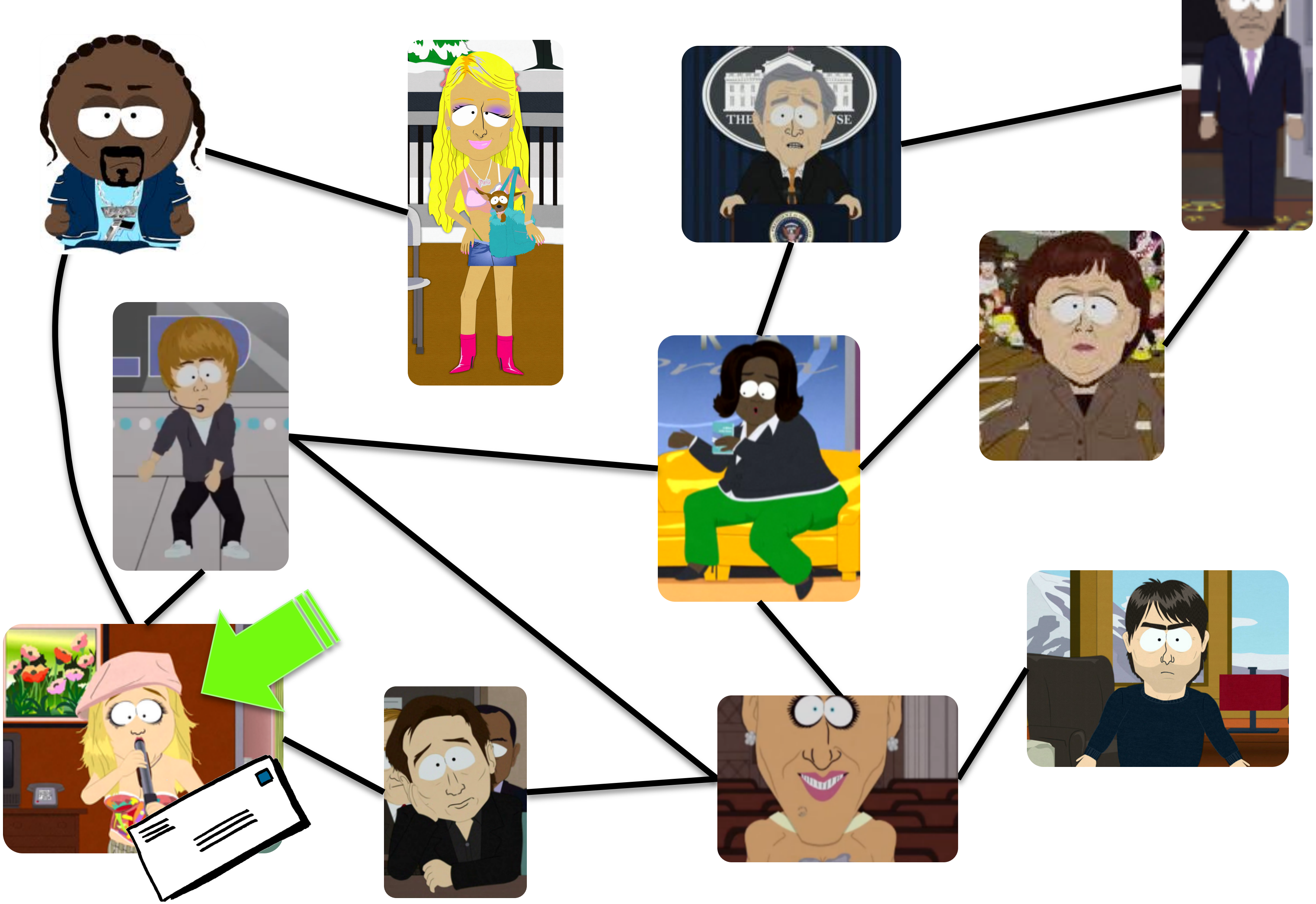
The social network of celebrities



The social network of celebrities



The social network of celebrities



The notion of the small world phenomenon

- Milgram's experiment form the 1960's
  - Given: Social network
  - Idea: How many jumps are needed to reach a random person?
  - Starters: Random people from Nebraska
  - Target: Person, who lived in Massachusetts and worked in Boston
  - Known: Basic information about the target
  - Rule: Send to a people known on first name basis
- Routing using only local information and simple facts
- Routing works!

  … and uses always very short paths

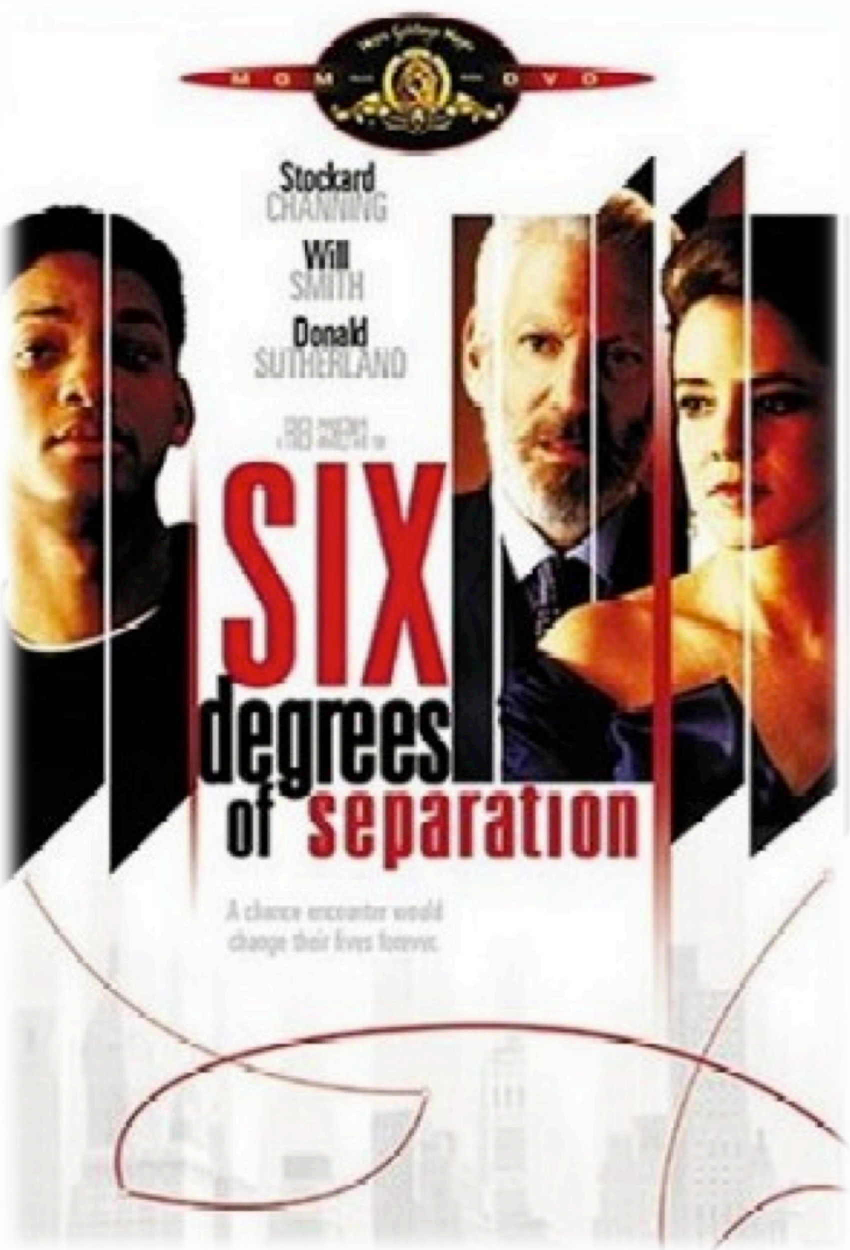
Results from Milgram’s experiment

- Chain: between 2 and 10 intermediate acquaintances
- Median: 5 intermediates
- Conclusion: Any person appeared to be reachable in just 6 jumps

Results from Milgram's experiment

- Chain: 2 to 10 intermediate acquaintances
- Median: 5 intermediates
- Conclusion: Any person appeared to be reachable in just 6 jumps

→ "Six degrees of separation"


MGM DVD
Stockard CHANNING
Will SMITH
Donald SUTHERLAND
SIX degrees of separation
A chance encounter would change their lives forever.

Six Degrees of Kevin Bacon

The Oracle of Bacon:

Six Degrees of Kevin Bacon

The Oracle of Bacon

Importance

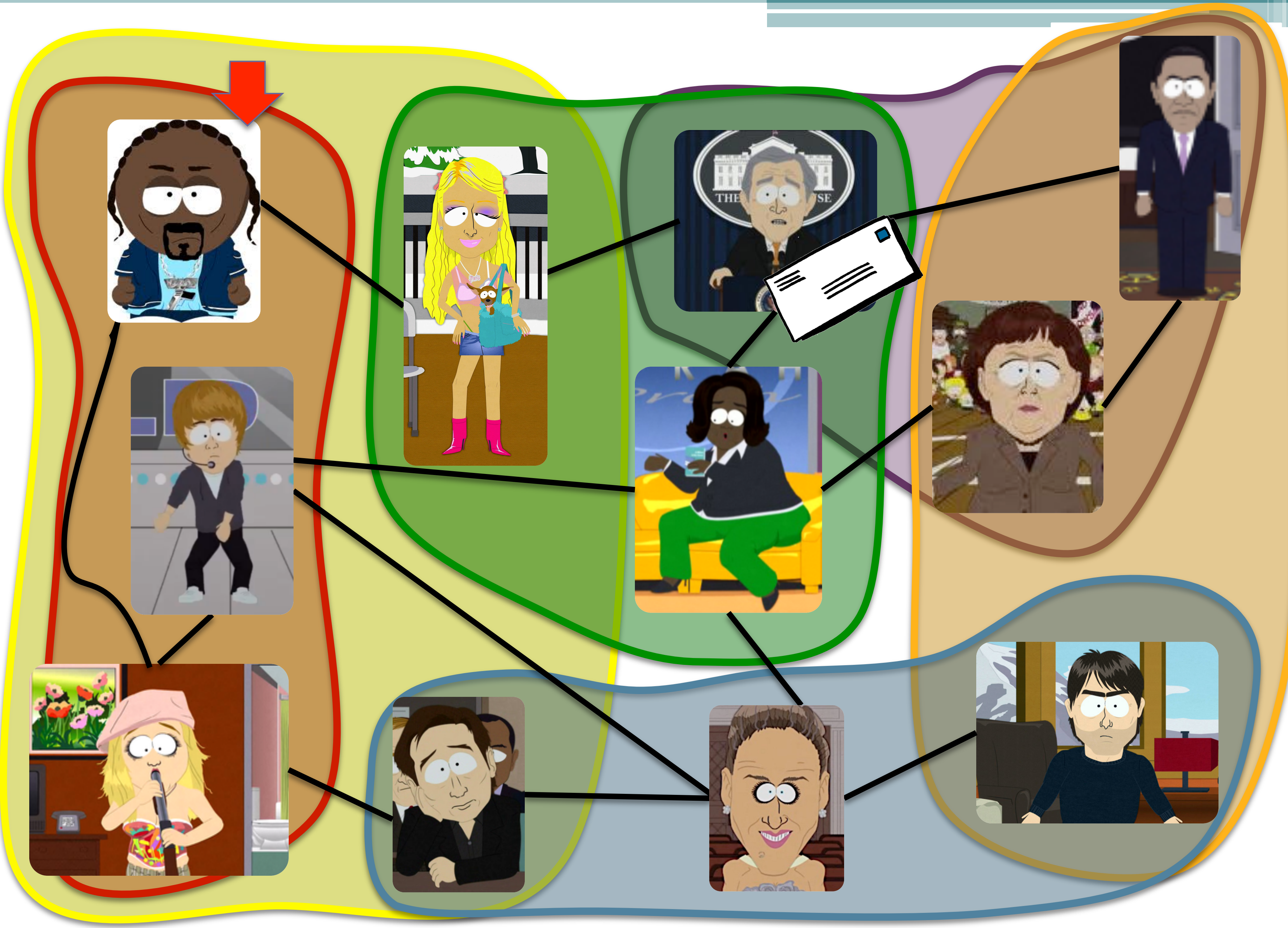
- Indicative of the underlying structure of modern social networks

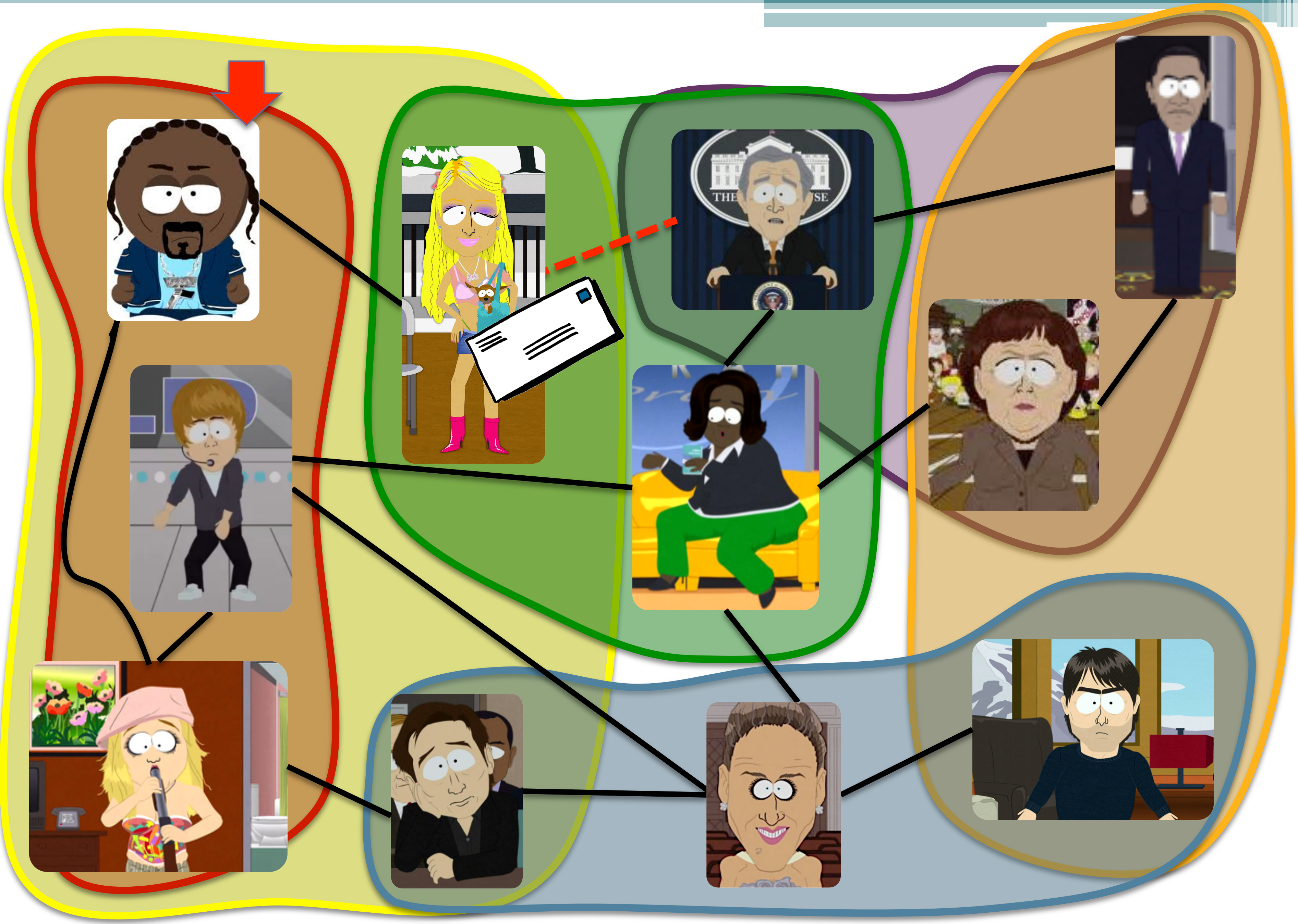
- In what other context:
  - infectious disease spreading
  - computer virus transmission

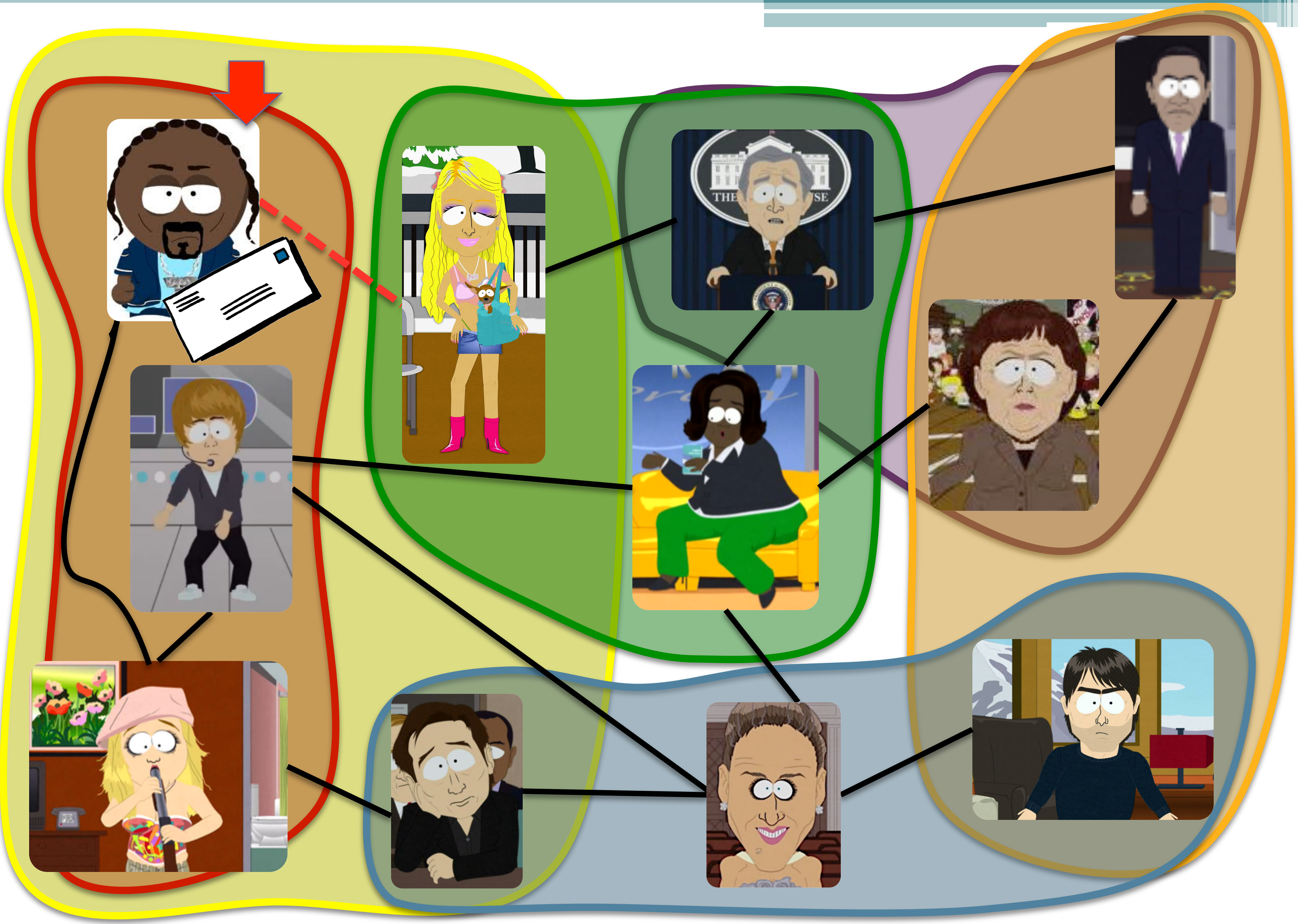
Observation from Milgram's experiment

- Decisions people take for selecting routes are overwhelmingly categorical in nature
- Categories are based on different factors like:
  - occupation, location, ethnicity

Consider a set of categories
… in the scenario of a social network of celebrities





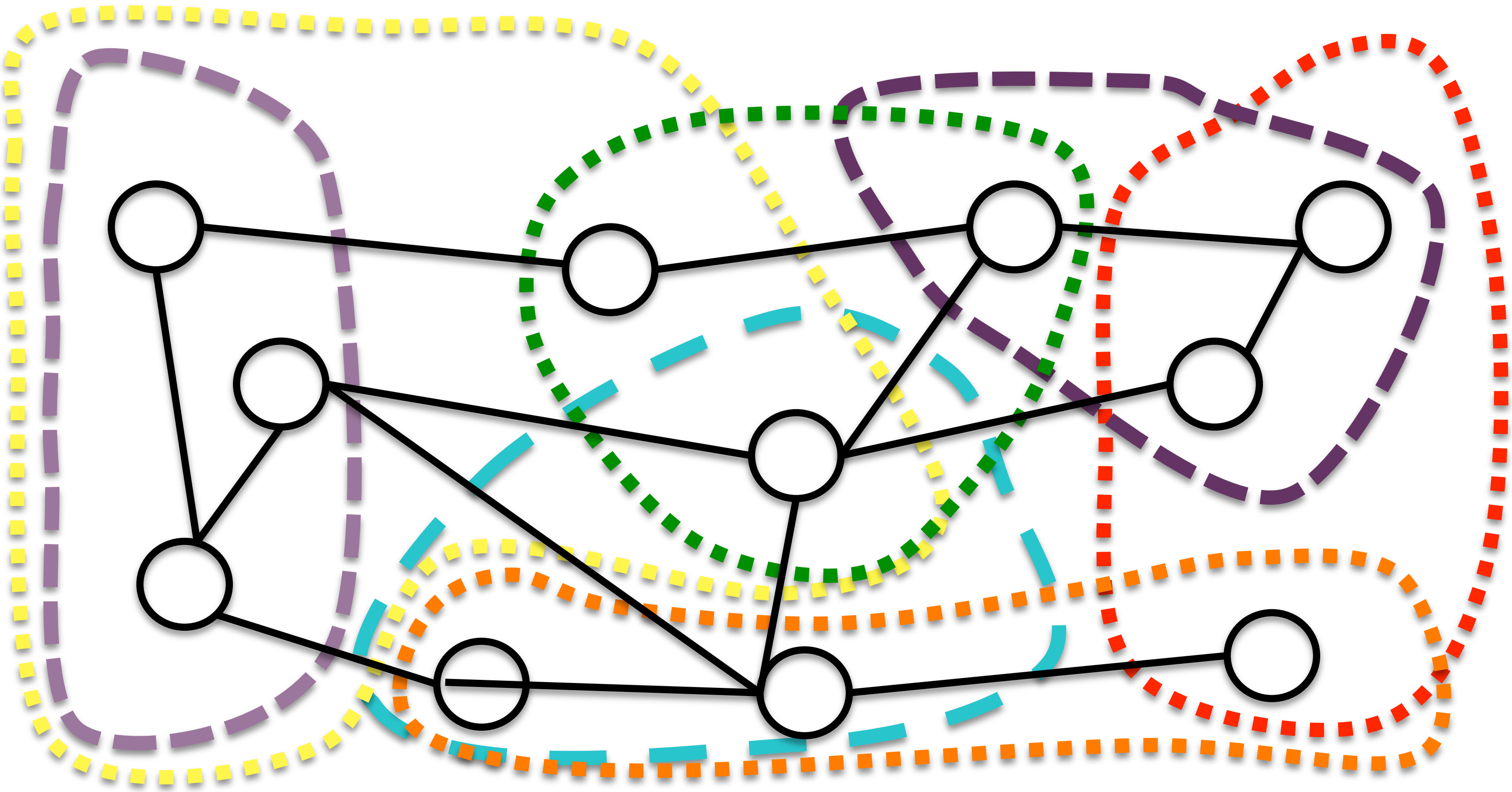


Part 2

Definitions and routing algorithm

Definitions:

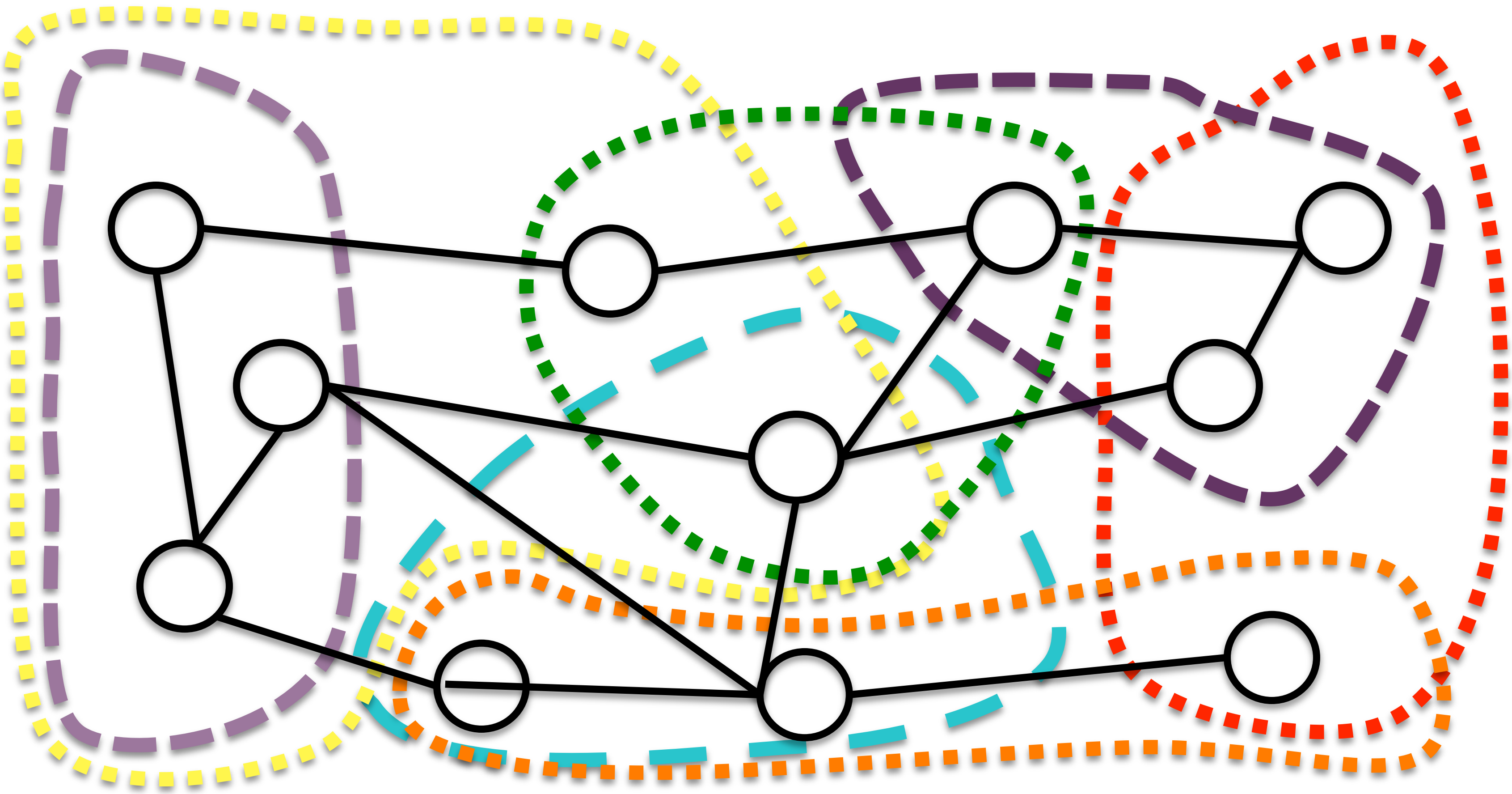
- Collection of categories: $S \subset 2^U$



Definitions

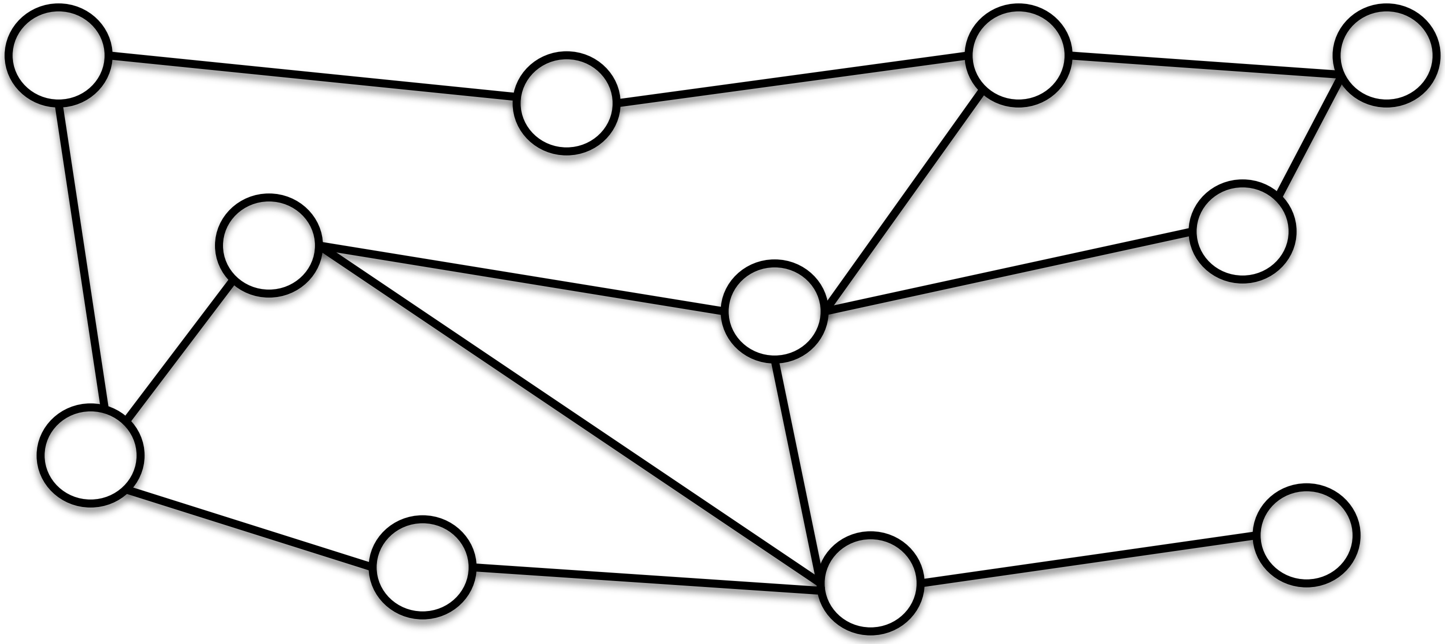
- Membership dimension: $\text{memdim}(S) = \max_{u \in U} |\text{cat}(u)|$

where $\text{cat}(u) = \{C \in S \mid u \in C\}$ is the set of categories of a node u



Definitions

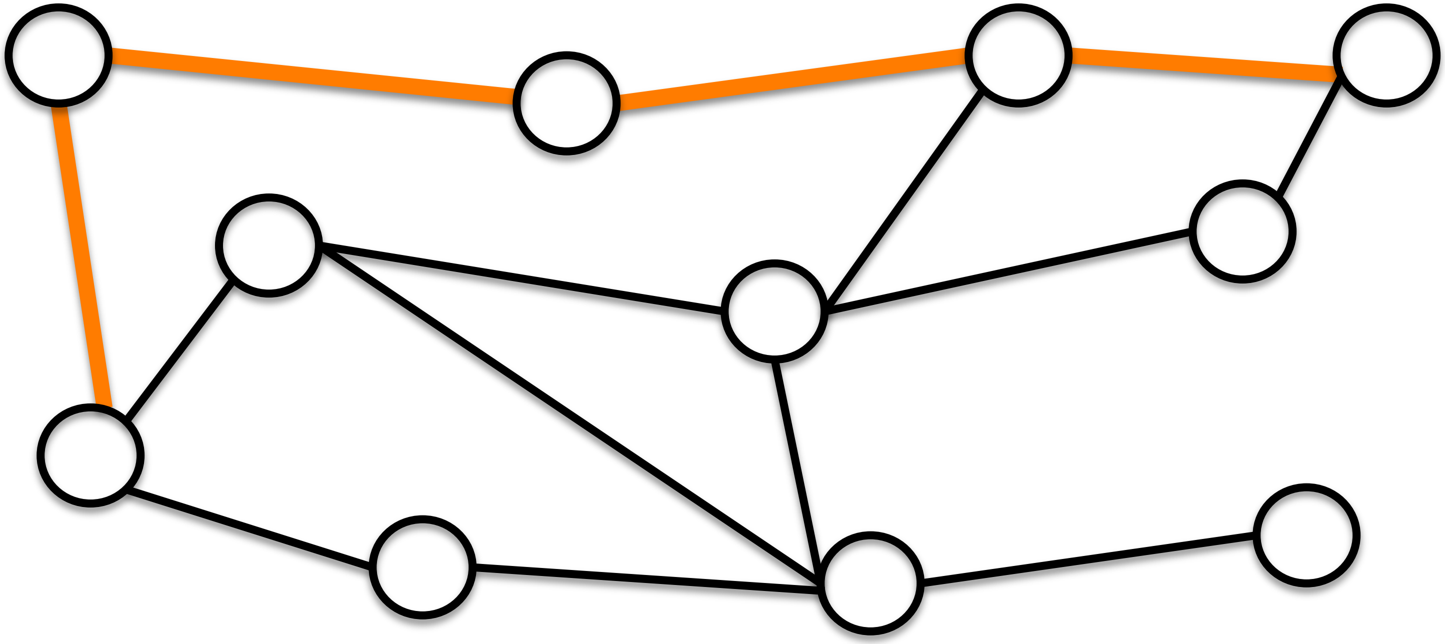
- Maximum length of any shortest path:
  $diam(G)=max_{s,t \in U} sp(s,t)$



Definitions

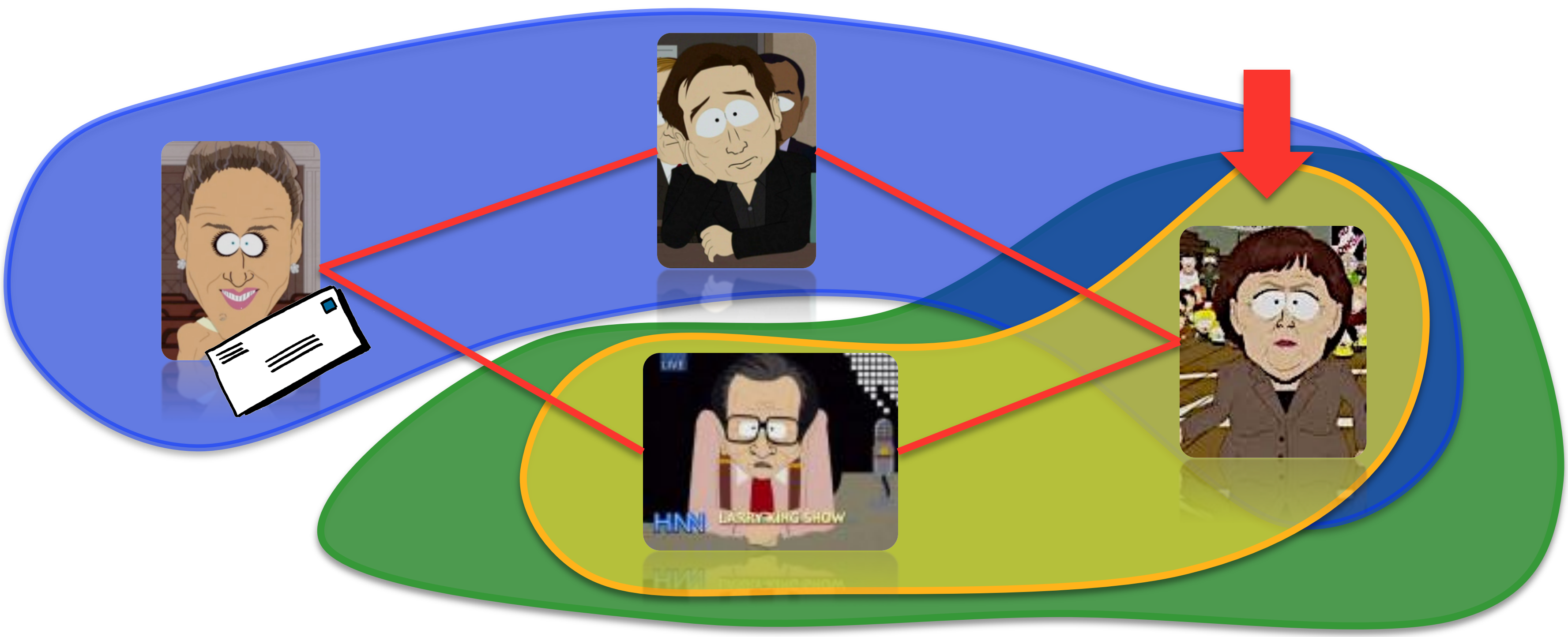
- Maximum length of any shortest path

  $\text{diam}(G)=\max_{s,t \in U} sp(s,t)$



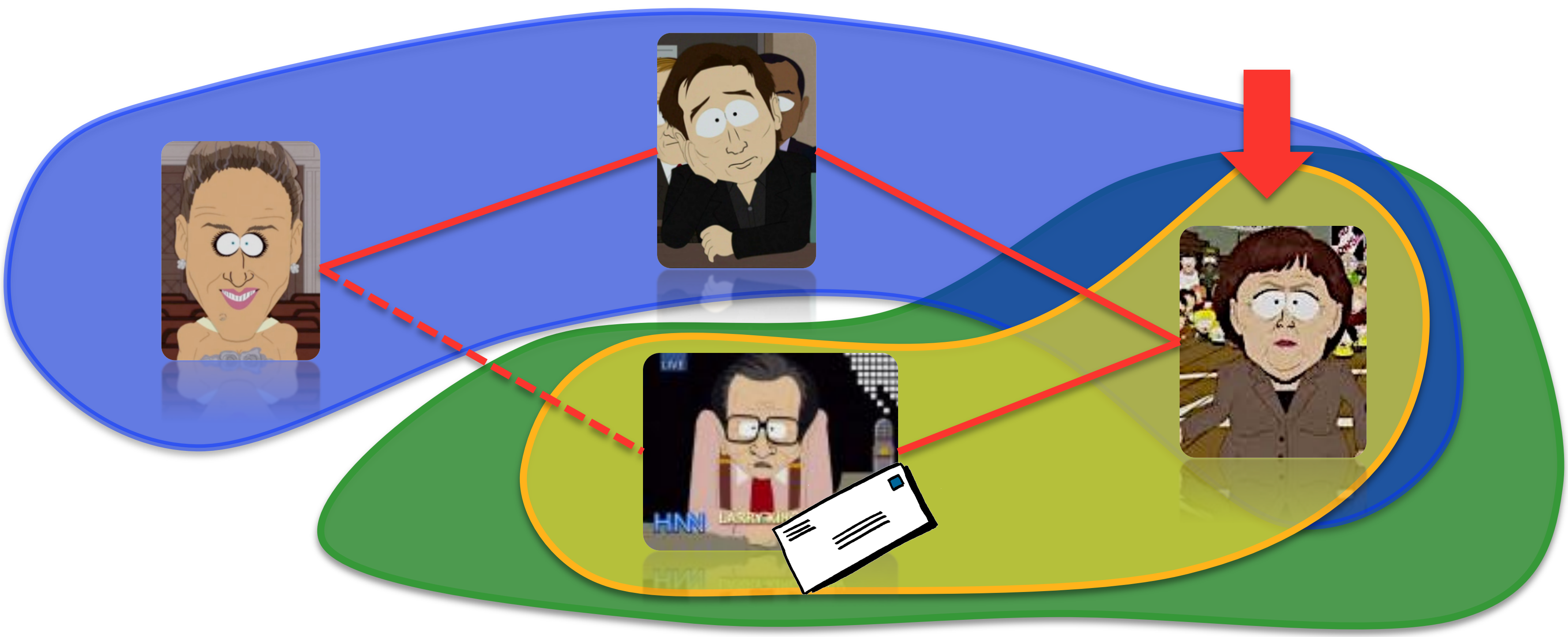
Greedy routing algorithm

- The category-based distance function: $d(s,t) = |cat(t) \setminus cat(s)|$
- Sending from u to w: forward to a neighbor v that is closer to w than u: $d(v,w) < d(u,w)$



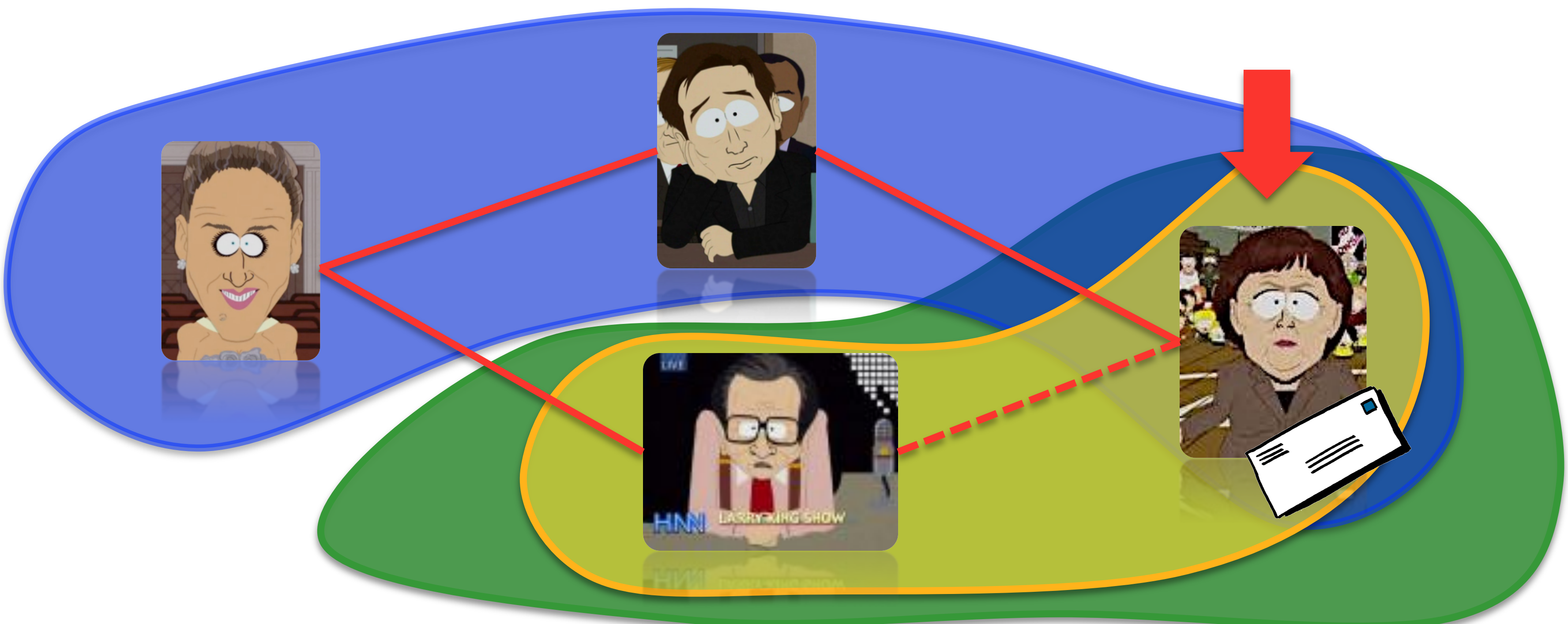
Greedy routing algorithm

- The category-based distance function: $d(s,t) = |cat(t) \setminus cat(s)|$
- Sending from u to w: forward to a neighbor v that is closer to w than u: $d(v,w) < d(u,w)$



Greedy routing algorithm

- The category-based distance function: $d(s,t) = |cat(t) \setminus cat(s)|$
- Sending from u to w: forward to a neighbor v that is closer to w than u: $d(v,w) < d(u,w)$



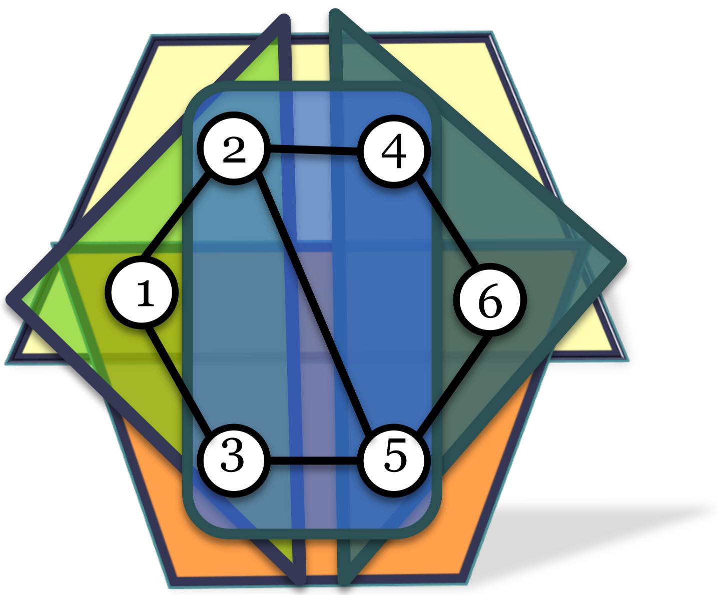
Properties of successful routing

- Internally connected:

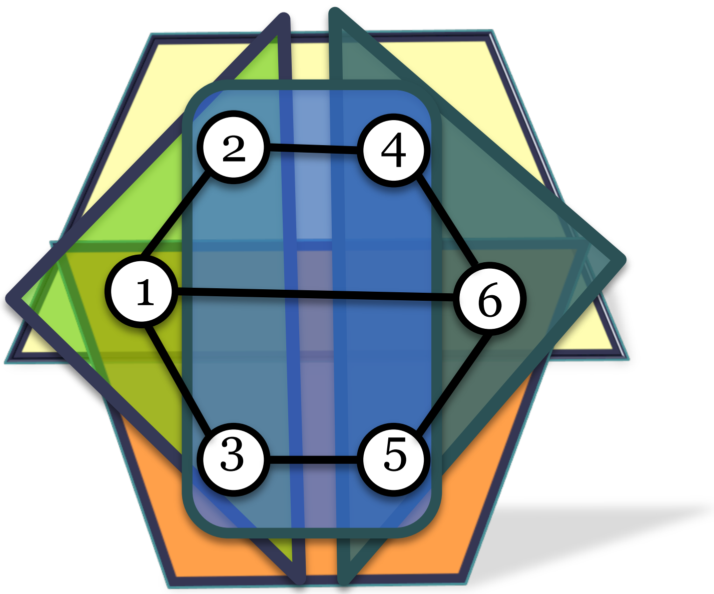
  (G, S) is internally connected, if for each C $\in$ S, G restricted to C is internally connected

- Shattered: A pair (G, S) is shattered if, for all $s, t \in U, s \neq t$, there is a neighbor u of s and a set $C \in S$ such that C contains u and t, but not s.

Examples



Internally connected
&
Not shattered



Shattered
&
Not internally connected

Shattered

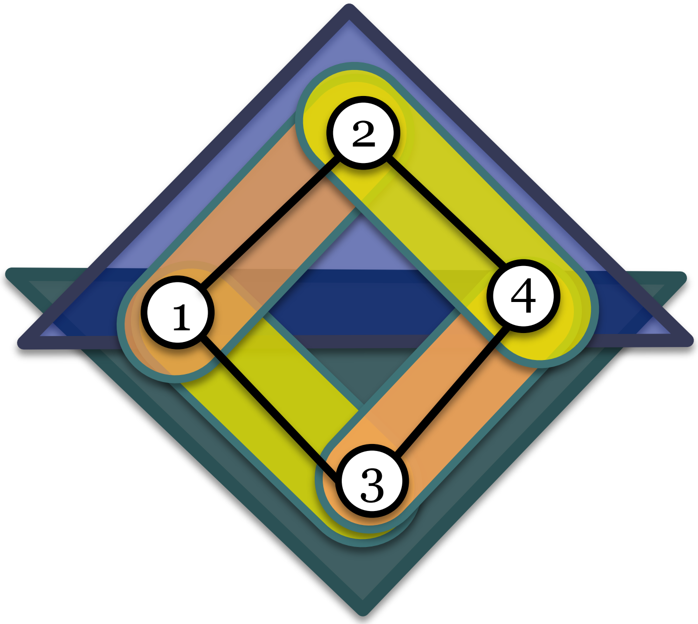
- In order for someone to advance a letter to a target, there must be an acquaintance that shares additional interests with the target.

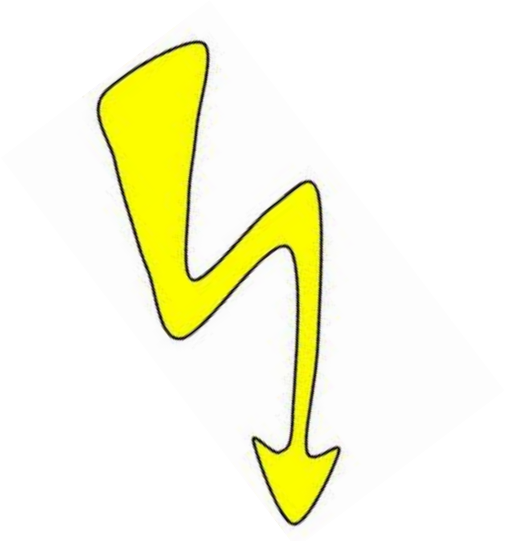
- **Lemma 1**: If (G, S) is not shattered, Routing fails.

What about routing in trees?

Lemma 2:
If G is a tree, and (G, S) is internally connected and shattered, then Routing is guaranteed to work.

- Not enough for arbitrary connected graphs





Summary of the definitions

- memdim(S)= $\max_{u \in U} |cat(u)|$
- $diam(G) = \max_{s,t \in U} sp(s,t)$
- Routing: forward to a neighbor v that: $d(v,w) < d(u,w)$
- Internally connected
- Shattered

Part 3

Building categories

Lower and upper bounds

If G and S are a graph and a category system such that Routing works:

- memdim(S) ≥ diam(G)
- $\text{memdim(S)} = O((\text{diam}(G) + \log n)^2)$

Lower bound of the cognitive load

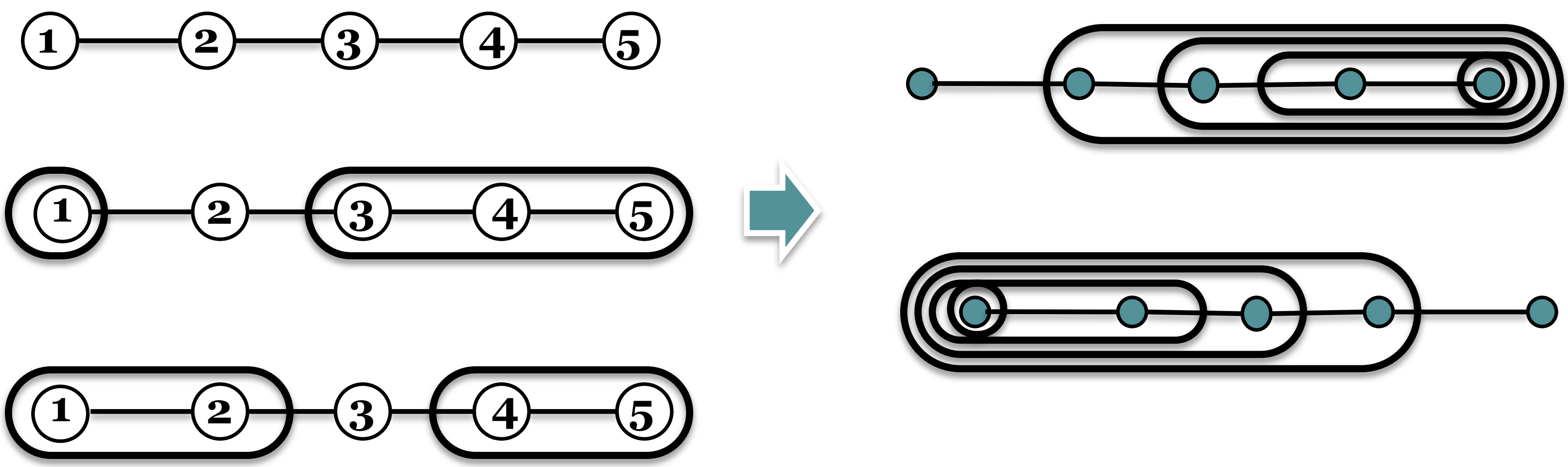
- Lemma 3:

If (G,S) be a graph and a category system, respectively, such that Routing works for G and S. Then $\text{memdim}(S) \geq \text{diam}(G)$

Routing in a graph as path

- Lemma 4:

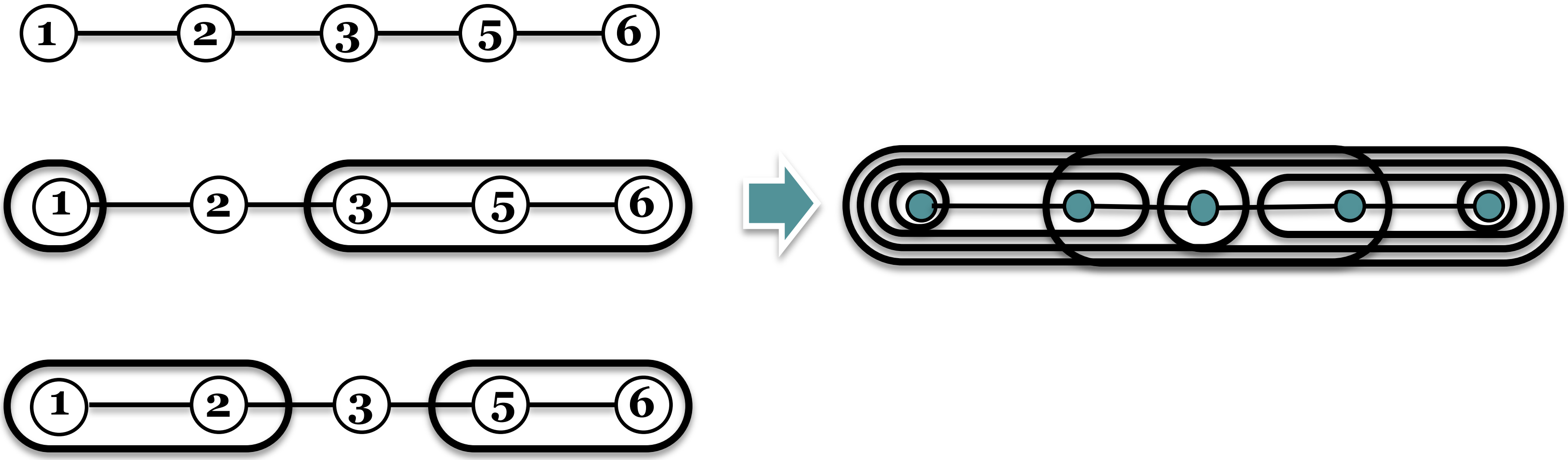
If G is a path, then there exists an S s.t. (G, S) is shattered and internally connected with memdim(S) = diam(G)



Routing in a graph as path

- Lemma 4:

If G is a path, then there exists an S s.t. (G, S) is shattered and internally connected with memdim(S) = diam(G)



Routing in a binary tree

- Lemma 5:

If G is a binary tree, then there exists an S s.t. (G,S) is shattered and internally connected with: $\text{memdim}(S) = O(\text{diam}^2(G))$

Why $\text{memdim}(S) = O(\text{diam}^2(G))$?

For $v \in U$: $u \in \text{ancestors}(v)$

$\Rightarrow v \in S_u$ and v belongs to O(height(u)) sets of L_u and R_u

$\Rightarrow$ v belongs to $O(\sum_{u \in \text{ancestor}(v)} \text{height}(u))$ sets

$\Rightarrow O(\text{diam}^2(G))$

Converting a n-node rooted tree

- Lemma 6:

Let T be an n-node rooted tree with height h. We can embed T into a binary tree such that the ancestor-descendant relationship is preserved, and the resulting tree has a height $O(h + \log n)$

Upper bound of the cognitive load

Theorem:

If G is connected, there exists S s.t. Routing works and memdim(S) $= O((\text{diam}(G) + \log n)^2)$

- Compute a low-diameter spanning tree T of G $\xrightarrow{\text{BFS}}$ $\text{diam}(T) \leq 2\text{diam}(G)$
- Arbitrary root T and embed T into a binary B

  with height $O(\text{diam}(T) + \log n)$, by Lemma 6
- $\text{diam}(B) = O(\text{diam}(T) + \log n)$
- By Lemma 5 $\Rightarrow$ $\text{memdim}(S_B) = O((\text{diam}(T) + \log n)^2)$
- From S_B to S_T and $\text{memdim}(S_T) \leq \text{memdim}(S_B) = O((\text{diam}(T) + \log n)^2)$

$$\Rightarrow \text{memdim}(S\ ) = O((\text{diam}(G) + \log n)^2)$$

Summary

- Arbitrary graph: memdim(S) $\geq$ diam(G)
- Path: memdim(S) = diam(G)
- Binary tree: $\text{memdim}(S) = O(\text{diam}^2(G))$
- From arbitrary to binary tree: height of $O(h + \log n)$
- Arbitrary graph: $\text{memdim(S)} = O((\text{diam}(G) + \log n)^2)$

Thank you!