Social Network Connectivity Algorithm

SoNCA

Edoardo Biagioni
University of Hawaii at Manoa

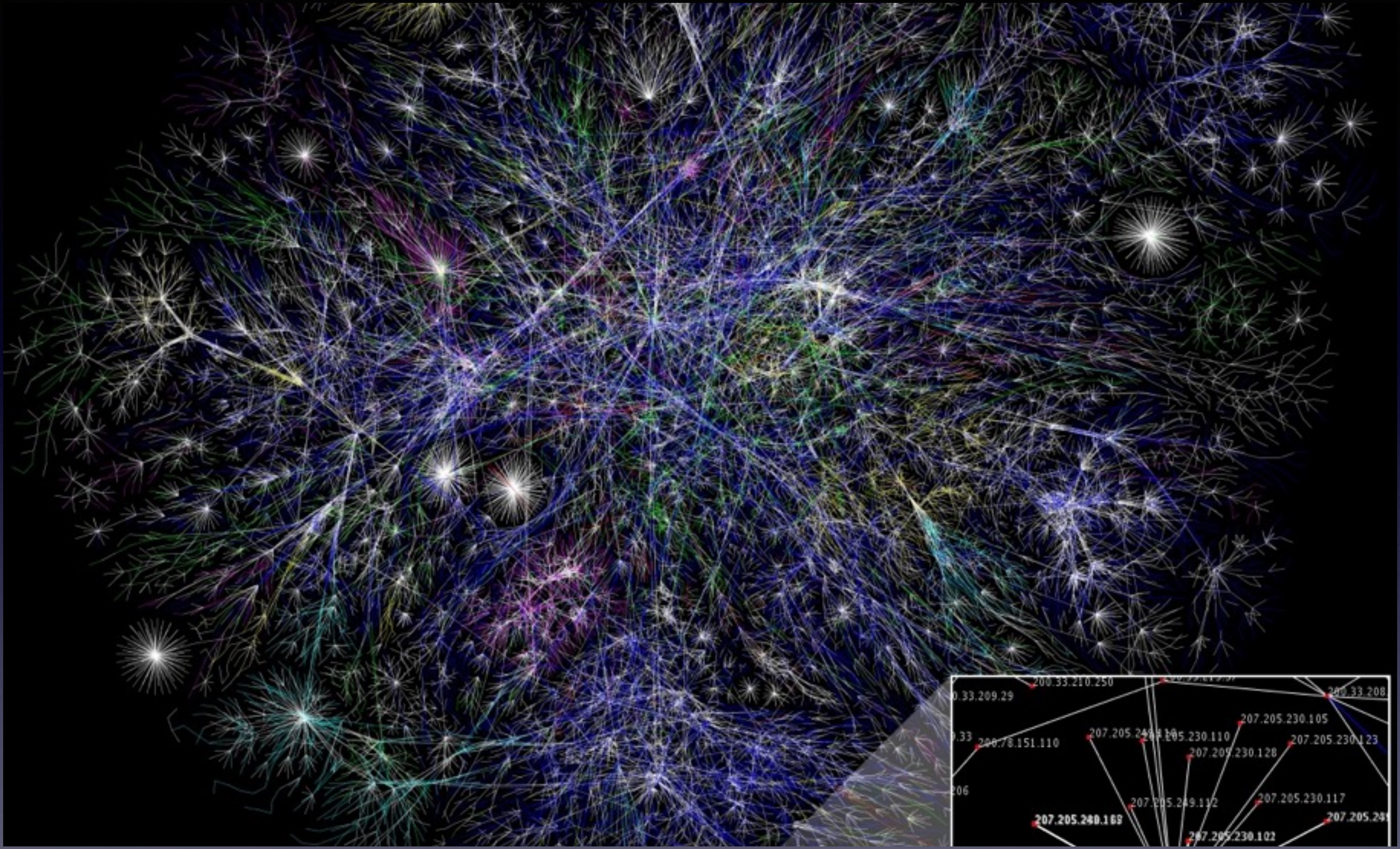
Why track social distance?

- When meeting for the first time
- To build a distributed social network (meet friends of friends)
- To know whom to favor when providing a free service
- For research and analysis
- ….

Rufino Uribe, CC-bySA 3.0

Social Distance

- My friends in set *f* are at distance *d = 1*
- Their friends are at *d = 2* from me, and in my f^2
  - the friends of my friends, that is, my **friends of friends**
- Their friends are at *d = 3* from me, and in my f^3
- If one of my contacts in *f* is in your f^3, our distance is *d = 4*



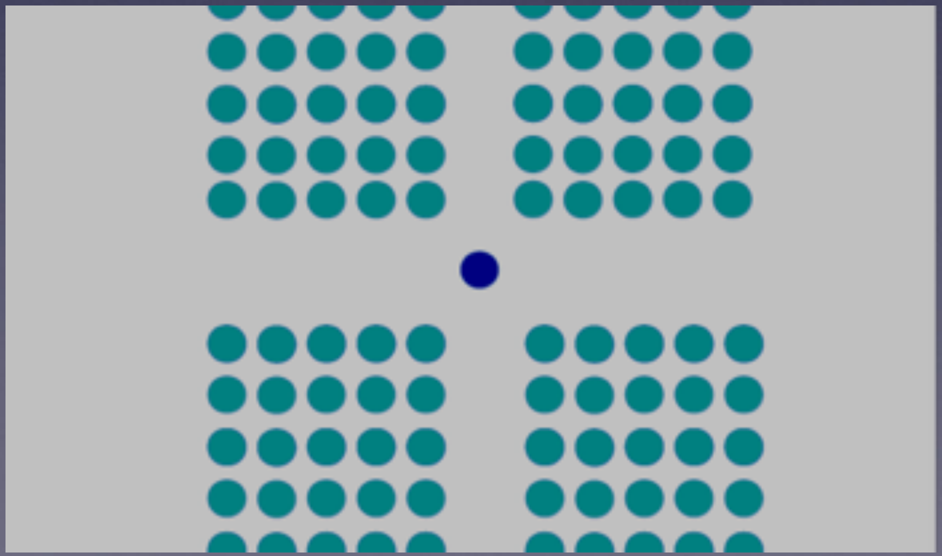
The Opte Project, CC-BY 2.5

Properties of Random Networks

- If every node has degree $|f|$,
  - and connections between nodes are random
- about $|f|^2$ nodes will be reachable at distance 2
  - and be in the node's f^2 set

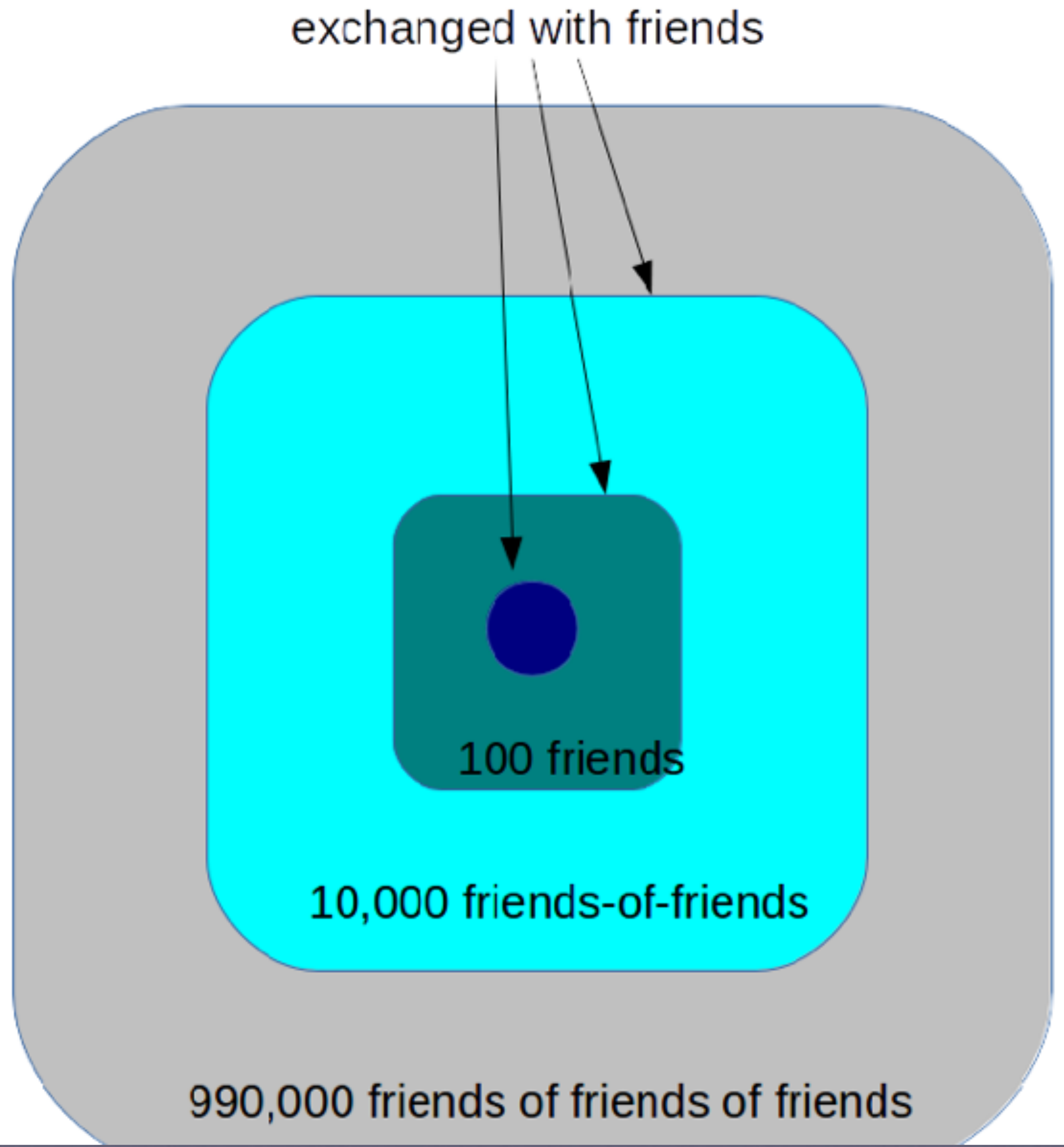
Distributed Social Network

- Every person or device picks a random k-bit ID
  - e.g. k = 128
    - very low chance of collision (2^{-64})
- and keeps track of all its friends' IDs



Network Maintenance

- Whenever I meet a friend, I give them k_2 bits of the IDs of all my friends f, and k_3 bits of the IDs of my friends-of-friends f^2
    - $k > k_2 > k_3$, e.g. $k = 128$, $k_2 = 112$, $k_3 = 96$
- everyone gradually builds their f, f^2, and f^3 sets
    - information is always “nearly complete”
        - as long as meeting new acquaintances is less common than meeting old friends


exchanged with friends
100 friends
10,000 friends-of-friends
990,000 friends of friends of friends

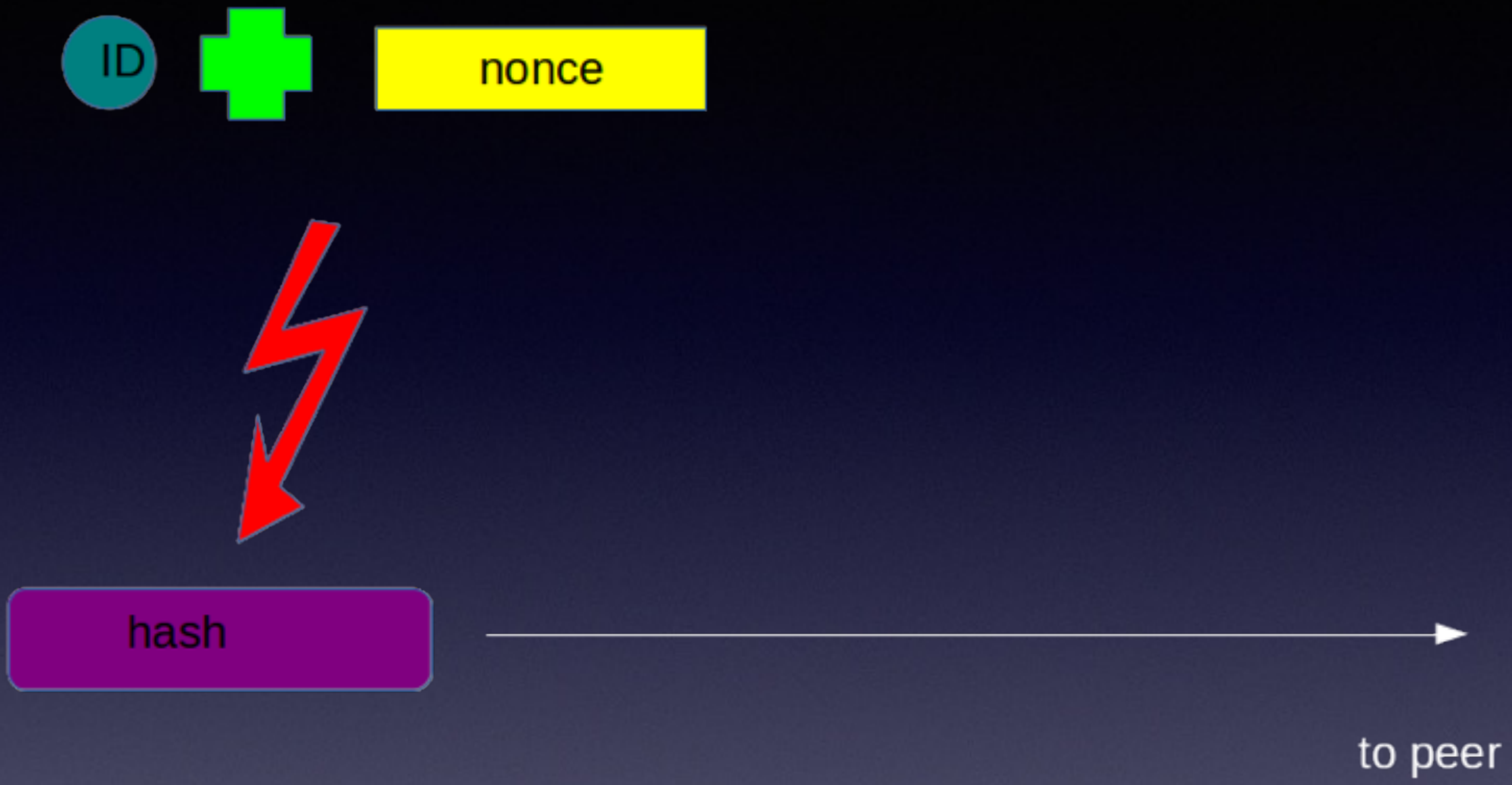
- Limit network size to a reasonable number, e.g. no more than 1 million IDs
  - most IDs will be in f^3, so 12MB of storage
  - if $|f| >= 100$ contacts, $|f3| >= 1$ million
- It is likely that two devices each with 1 million random IDs, will share at least one ID
  - if the world has less than 1 trillion IDs
  - average social distance is 6 or less

Distributed Computation of Social Distance — oversimplified

- You give you all the IDs in your *f*, *f*^{2}, and *f*^{3}
- I figure out which is closest to me
- if necessary, I give you all my IDs
- problem: now you have all my IDs and can pretend to be close to my friends

Social Network Computation Algorithm, SoNCA

- We agree on a nonce that includes the current time and date (to prevent replay attacks) and information about me and you
- You hash each ID concatenated with the nonce
    - and send me the hashes
- Now I can verify your distance
    - but I do not have any of your IDs, so cannot use your information with someone else



ID

nonce

hash

to peer

SoNCA optimization

- 96 bits for each of a million IDs means the IDs are very sparse
- we can begin by exchanging a sparse bitmap to indicate where we have IDs
- then only hash and exchange IDs where the bitmap shows we both have a potential match

Effectiveness of the optimization

- If both sides are honest, reduces the amount of hashing and data exchange
- Cheating brings no benefit:
  - evidence is provided by the hash, not the bitmap
- there is little harm in adding additional, fictional bits to the bitmap
  - the other party may compute more hashes than necessary

SoNCA summary

- Keep track of up to 1 million IDs, with fewer bits for IDs with greater social distance
- Exchange hashes of IDs
    - hashed with unique nonce
- Use bitmaps to reduce number of hashes

Using SoNCA: AllNet

- Distributed P2P network for interpersonal communications
  - “my cellphone talks directly to your cellphone”
- Forward messages for others (multihop)
  - better to limit resource consumption
  - so, prioritize messages to and from friends!
  - to a lesser extent, from friends’ friends

The world is not random

- But it is random enough
- social connections obey power laws:
    - some of my friends are not your friends
    - some of my friends are “far away”
        - the network has short paths to anywhere

Evidence that real social networks behave like random networks

- social distance <= 6
    - measured in the 1960s by Milgram using postal mail
        - “six degrees of separation”
- more recent work at Facebook shows distance of less than 5

Other fruitless ways of cheating

- I pick your ID as my ID
  - hard to do (unless you are my friend), and
  - doesn't give you k=128 bits of my friends' IDs
- I make up random IDs for my friends
  - very unlikely you will match 96 bits of my f^3