

Who Contributes to the Knowledge Sharing Economy?

Arthi Ramachandran, Augustin Chaintreau

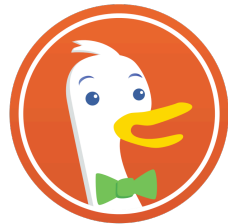
Columbia University

Nov 2, 2015

Knowledge Sharing Economy

Google

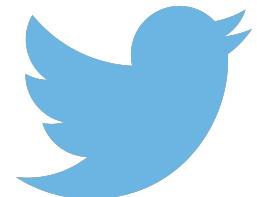
bing



DuckDuckGo

Search

vs

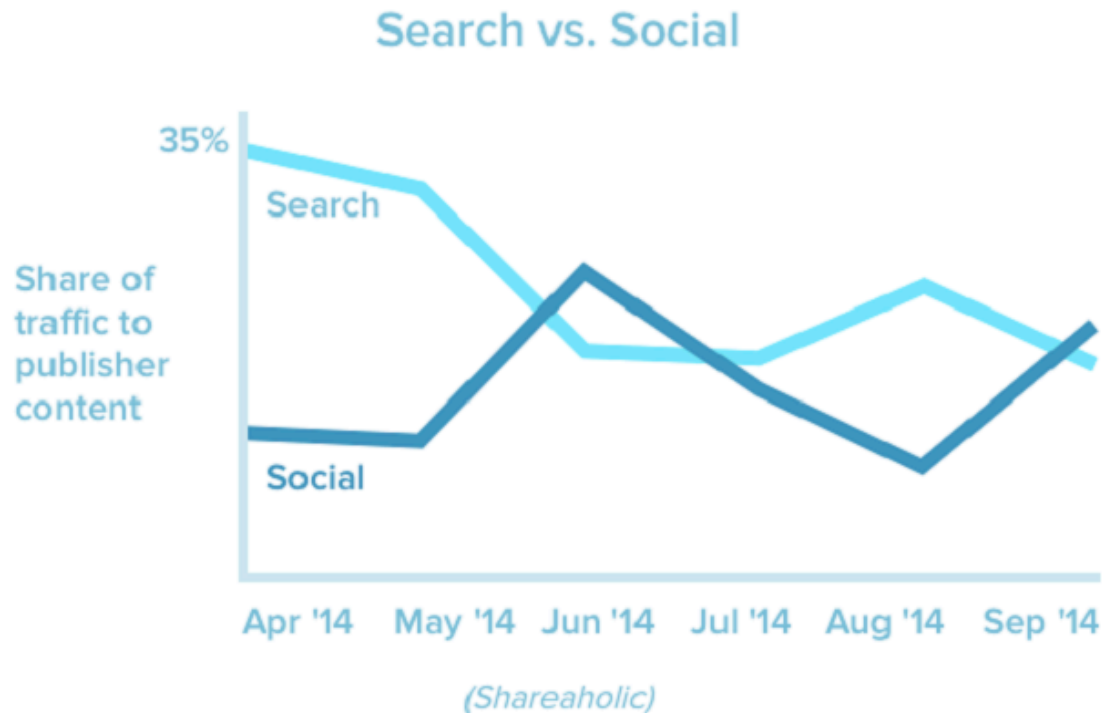


LinkedIn

Social

Who brings more people to websites?

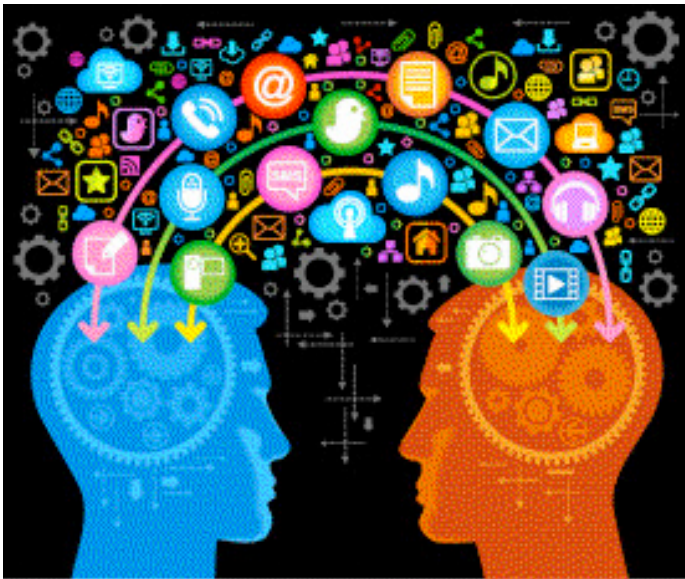
Knowledge Sharing Economy



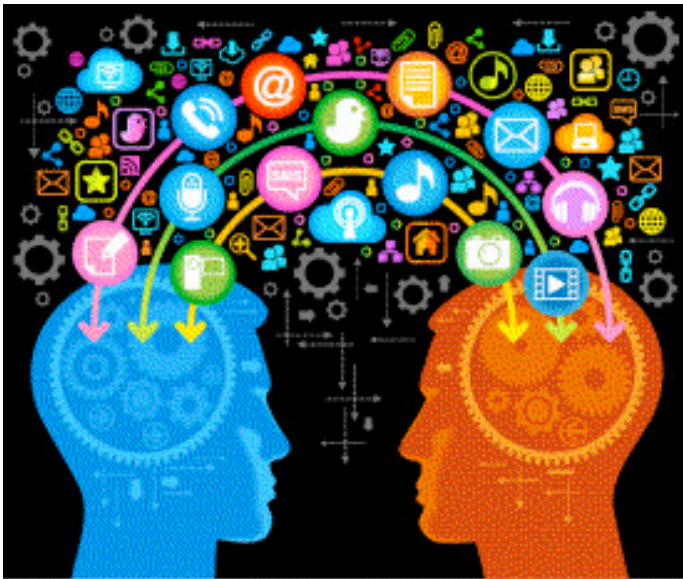
- Social is increasing in the share of referrals

Knowledge Sharing Economy

- Similar to the sharing economy (eg: uber, airbnb), where users benefit from other users' surplus



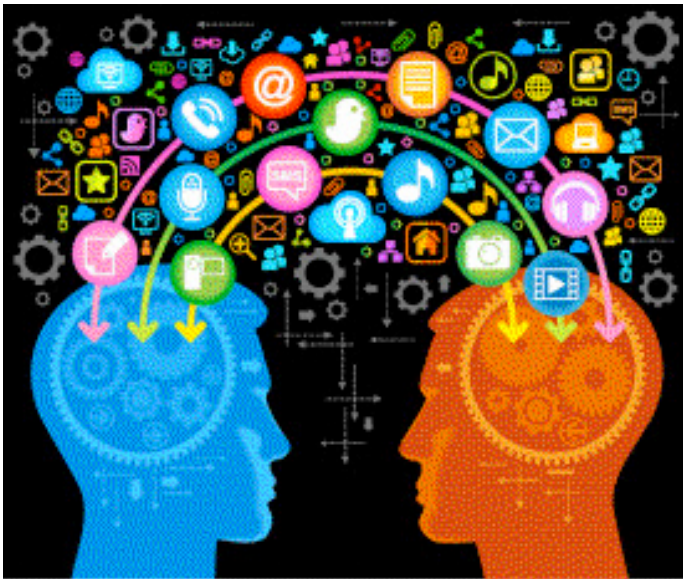
Knowledge Sharing Economy



- Similar to the sharing economy (eg: uber, airbnb), where users benefit from other users' surplus
- In the *Knowledge Sharing Economy*, other users benefit from users' information search

Knowledge Sharing Economy

- Information is
 - Financially important
 - In fact, intermediaries can earn a living by curating content ^{1,2}
 - Used for decision making ^{3,4}
 - Eg: job hunting, voting, new products



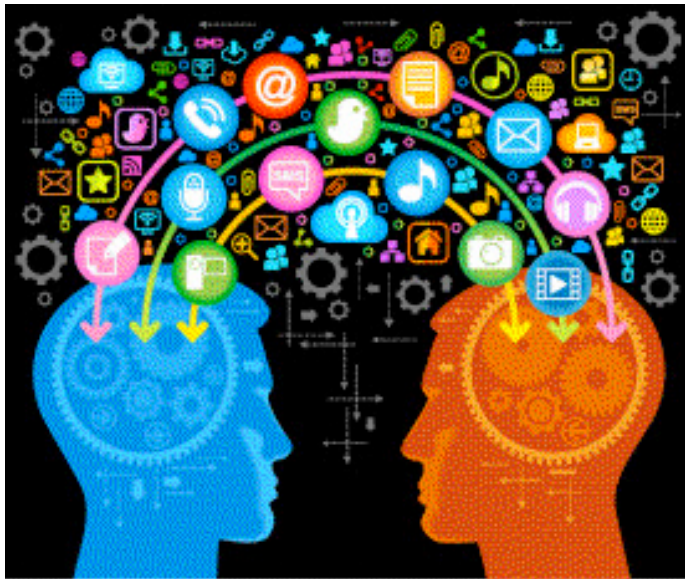
¹ M. Cha, F. Benevenuto, H. Haddadi, and K. Gummadi. The World of Connections and Information Flow in Twitter. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2012.

² S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. WWW 2011.

³ D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis) information in social networks. Games and Economic Behavior, 2010.

⁴ B. Golub and M. O. Jackson. Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics, 2010.

Knowledge Sharing Economy



- Information is
 - Financially important
 - In fact, intermediaries can earn a living by curating content ^{1,2}
 - Used for decision making ^{3,4}
 - Eg: job hunting
- These analyze information acquisition as an economic rational process

¹ M. Cha, F. Benevenuto, H. Haddadi, and K. Gummadi. The World of Connections and Information Flow in Twitter. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2012.

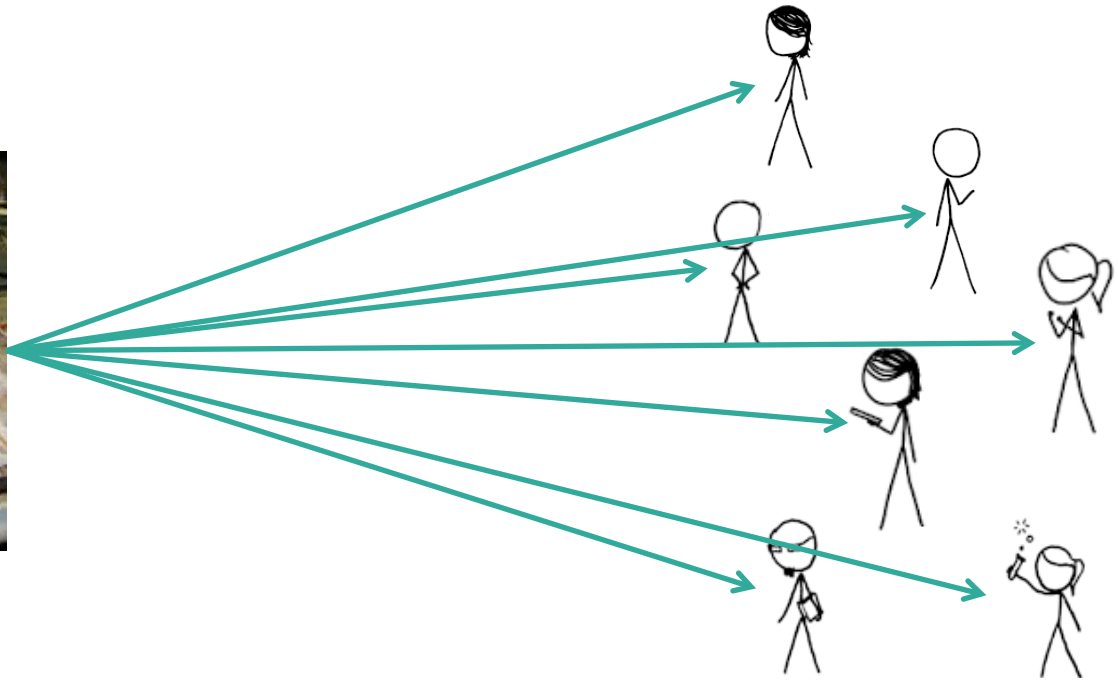
² S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. WWW 2011.

³ D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis) information in social networks. Games and Economic Behavior, 2010.

⁴ B. Golub and M. O. Jackson. Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics, 2010.

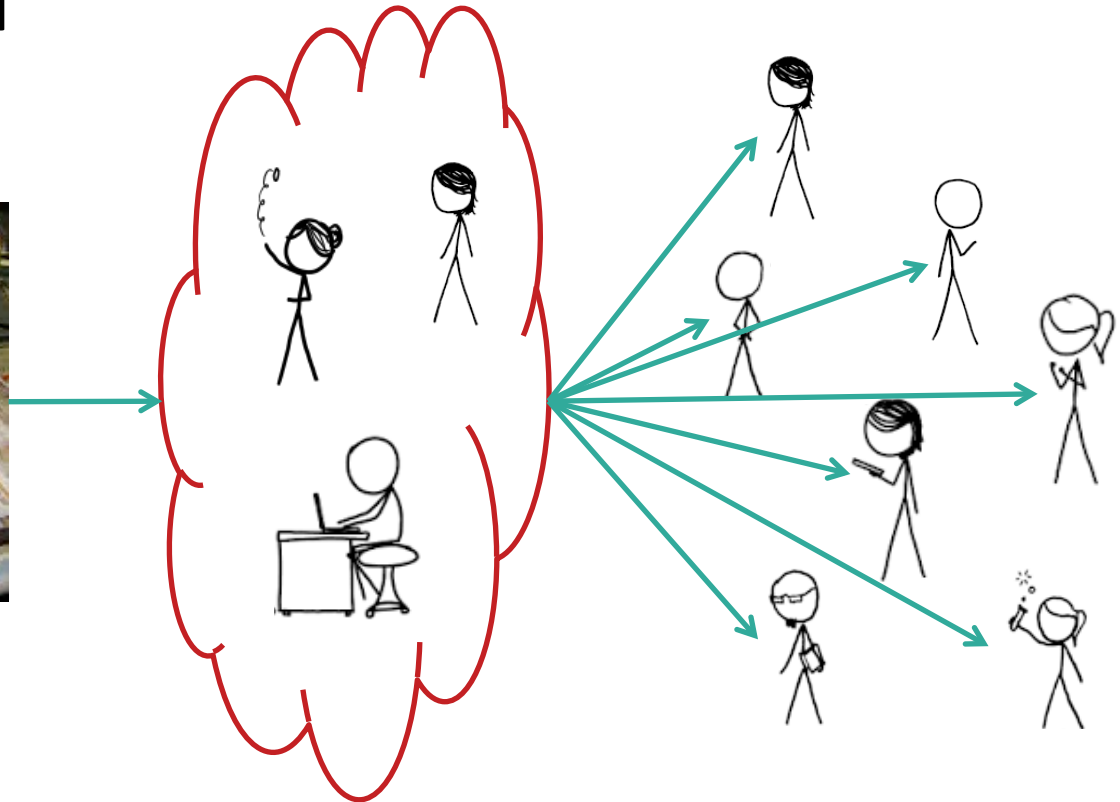
Models of content sharing

- Past: publisher -> reader model



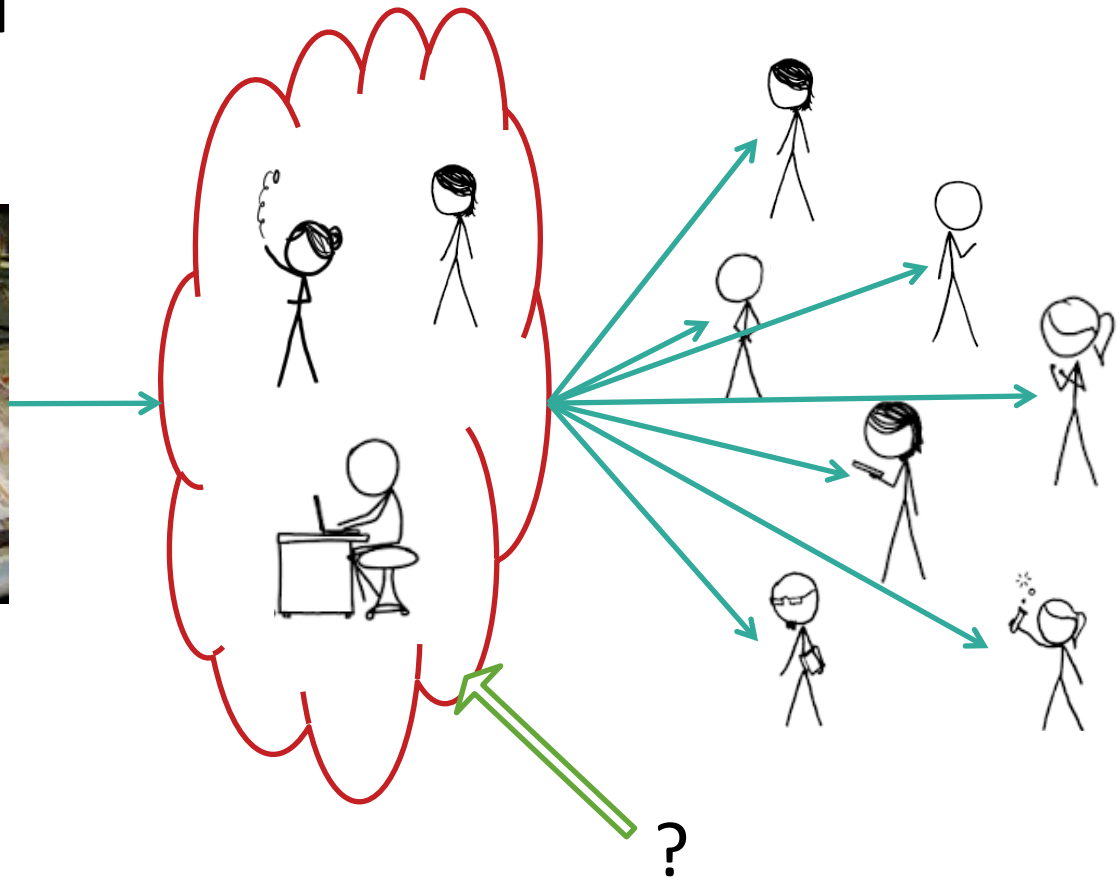
Models of content sharing

- With social networks: publisher -> curators -> reader model



Models of content sharing

- With social networks: publisher -> curators -> reader model



Types of sharing

- Where is the content coming from? Who looks for content to share on twitter?

Types of sharing

- Where is the content coming from? Who looks for content to share on twitter?
- Heterogeneous sharing
 - Content that many users find and share



Types of sharing

- Where is the content coming from? Who looks for content to share on twitter?
- Heterogeneous sharing
 - Content that many users find and share
 - Content that is more specialized



Marc Andreessen
@pmarca



"17 of US's largest metro economies as separate nations would rank in the top 50 largest economies in the world." aei.org/publication/un...

1	New York-Newark-Jersey City, NY-NJ-PA	\$1,558	Australia	\$1,444
2	Los Angeles-Long Beach-Anaheim, CA	\$807	Netherlands	\$805
3	Chicago-Naperville-Elgin, IL-IN-WI	\$805	Nigeria	\$818
4	Houston-The Woodlands-Sugar Land, TX	\$525	Taiwan	\$530
5	Dallas-Fort Worth-Arlington, TX	\$584	Hong Kong	\$500
6	Washington-Arlington-Alexandria, DC-VA-MD-WV	\$471	Austria	\$437
7	San Francisco-Oakland-Hayward, CA	\$432	Iran	\$406
8	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	\$391	Italy	\$402
9	Boston-Cambridge-Newton, MA-NH	\$385	Colombia	\$365
10	Atlanta-Sandy Springs-Roswell, GA	\$325	Malaysia	\$327
11	Buffalo-Tonawanda-Buffalo, NY	\$301	Singapore	\$308
12	Miami-Fort Lauderdale-West Palm Beach, FL	\$299	Israel	\$304
13	Detroit-Warren-Dearborn, MI	\$257	Greece	\$250
14	Minneapolis-St. Paul-Bloomington, MN-WI	\$238	Portugal	\$230
15	Phoenix-Mesa-Gilbert, AZ	\$215	Iran	\$211
16	San Jose-Sunnyvale-Santa Clara, CA	\$214	Algeria	\$214
17	San Diego-Carlsbad, CA	\$207	Venezuela	\$200

Understanding America's ridiculously large \$17.4 trillion economy by compari...

The table above helps to put America's ridiculously large \$17.4 trillion economy (GDP in 2014) into perspective by comparing America's largest 20 metro economi...

aei.org

Outline

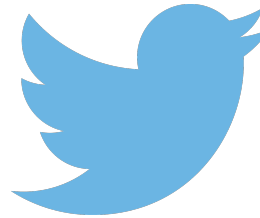
- Datasets
- Who are the Content Creators?
- Relationship of Content Lifespan and Concentration
- Model of Perishable Public Goods
- Equilibria and Specialization
- Conclusion

Outline

- Datasets
- Who are the Content Creators?
- Relationship of Content Lifespan and Concentration
- Model of Perishable Public Goods
- Equilibria and Specialization
- Conclusion

Datasets

- KAIST: Twitter posts over July 2009
 - Breadth – all twitter users and posts for a month
 - 8m unique users
 - 37m unique URLs
- NYT: Twitter posts containing URLs to nytimes.com for 1 week in Dec 2012
 - Depth – all twitter users receiving certain urls
 - 346k unique users
 - 70k unique URLs

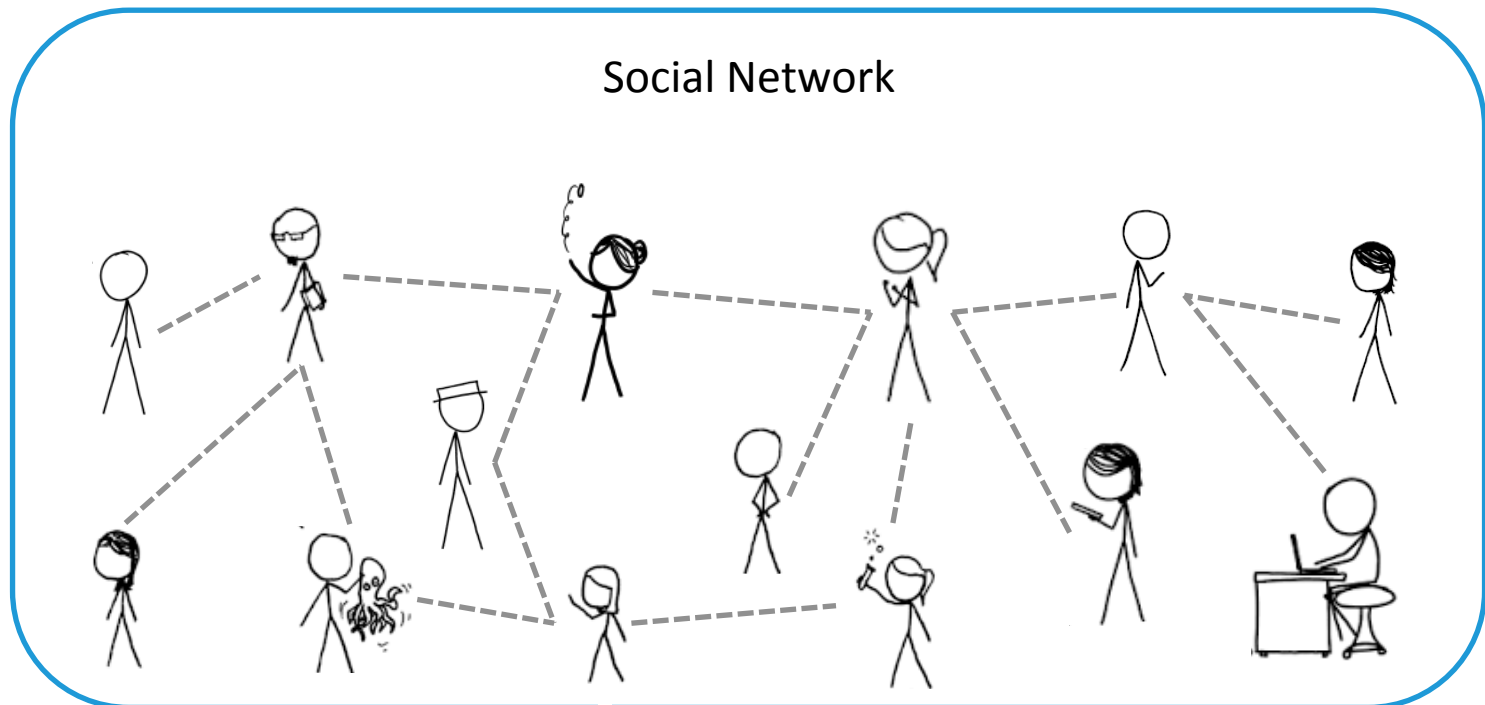


Outline

- Datasets
- Who are the Content Creators?
- Relationship of Content Lifespan and Concentration
- Model of Perishable Public Goods
- Equilibria and Specialization
- Conclusion

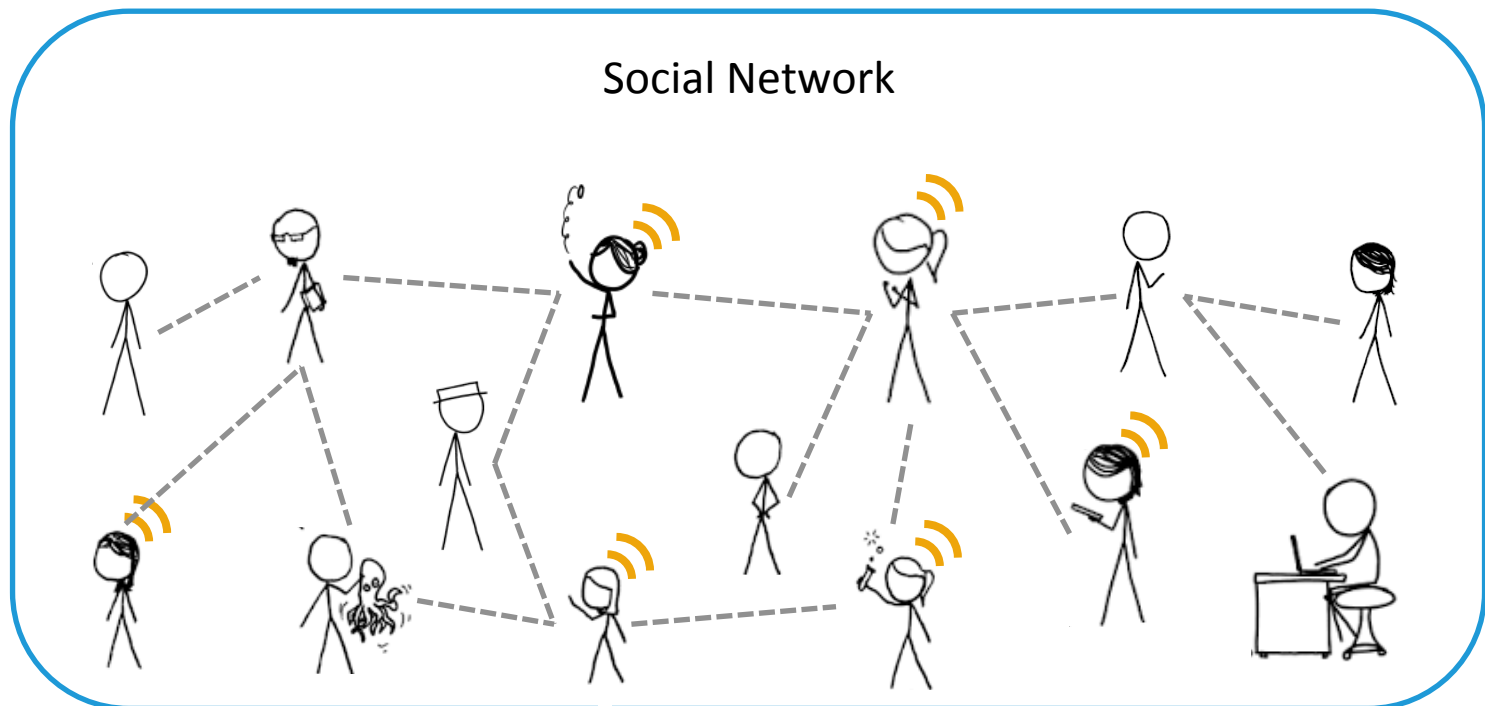
Who is Responsible for Creating Content?

- Network of users



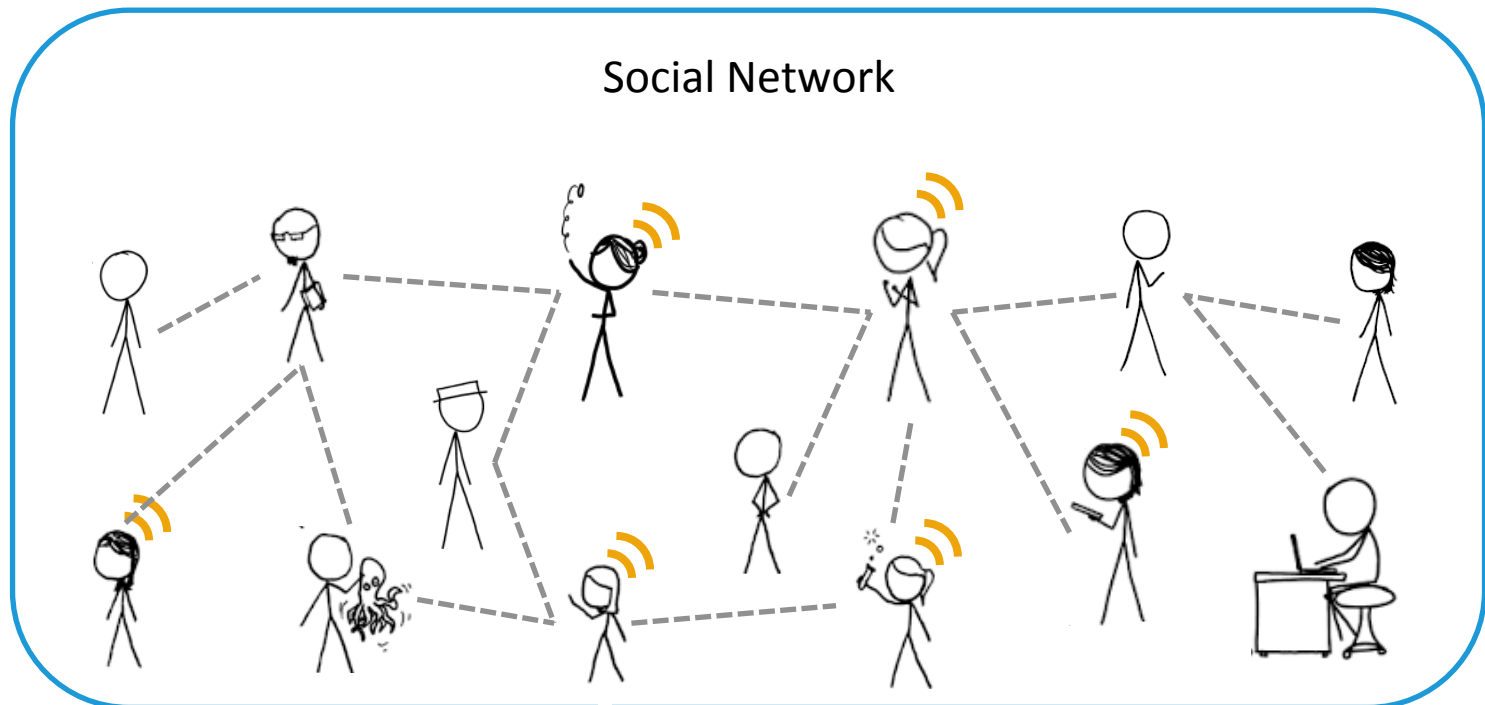
Who is Responsible for Creating Content?

- Network of users posting (📶) content



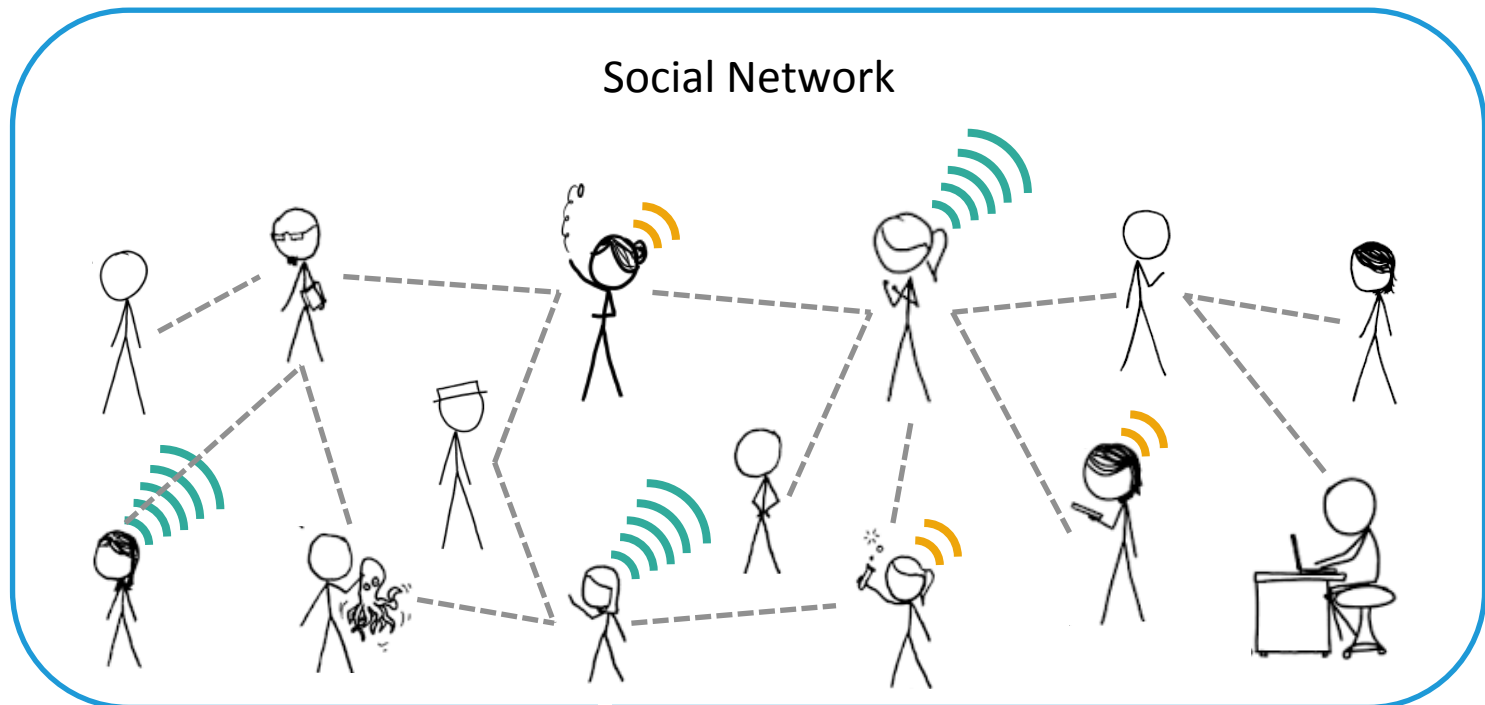
Who is Responsible for Creating Content?

- Different classes of user posts:
 - Anyone who posts



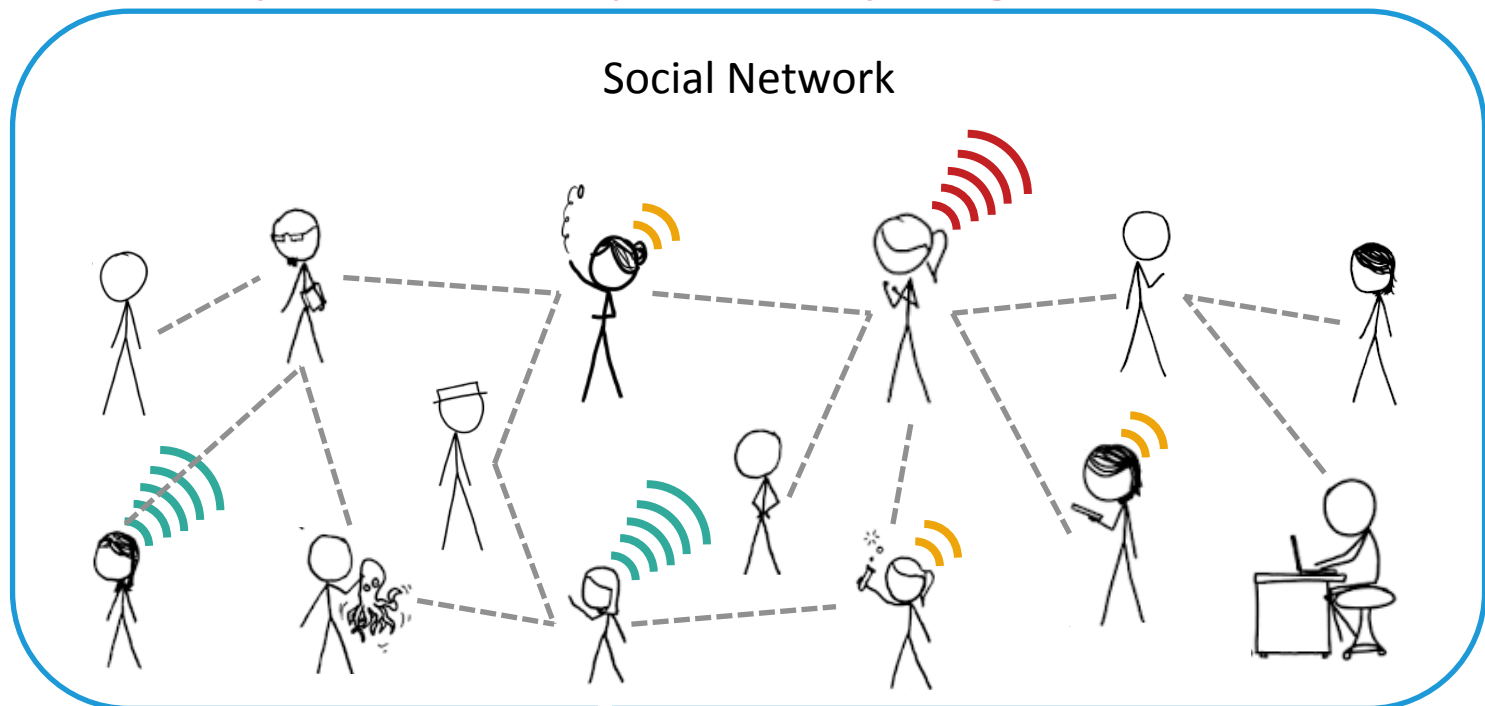
Who is Responsible for Creating Content?

- Different classes of user posts:
 - Anyone who posts
 - Locally the first to post



Who is Responsible for Creating Content?

- Different classes of user posts:
 - Anyone who posts
 - Locally the first to post
 - Globally the first to post (very original content)

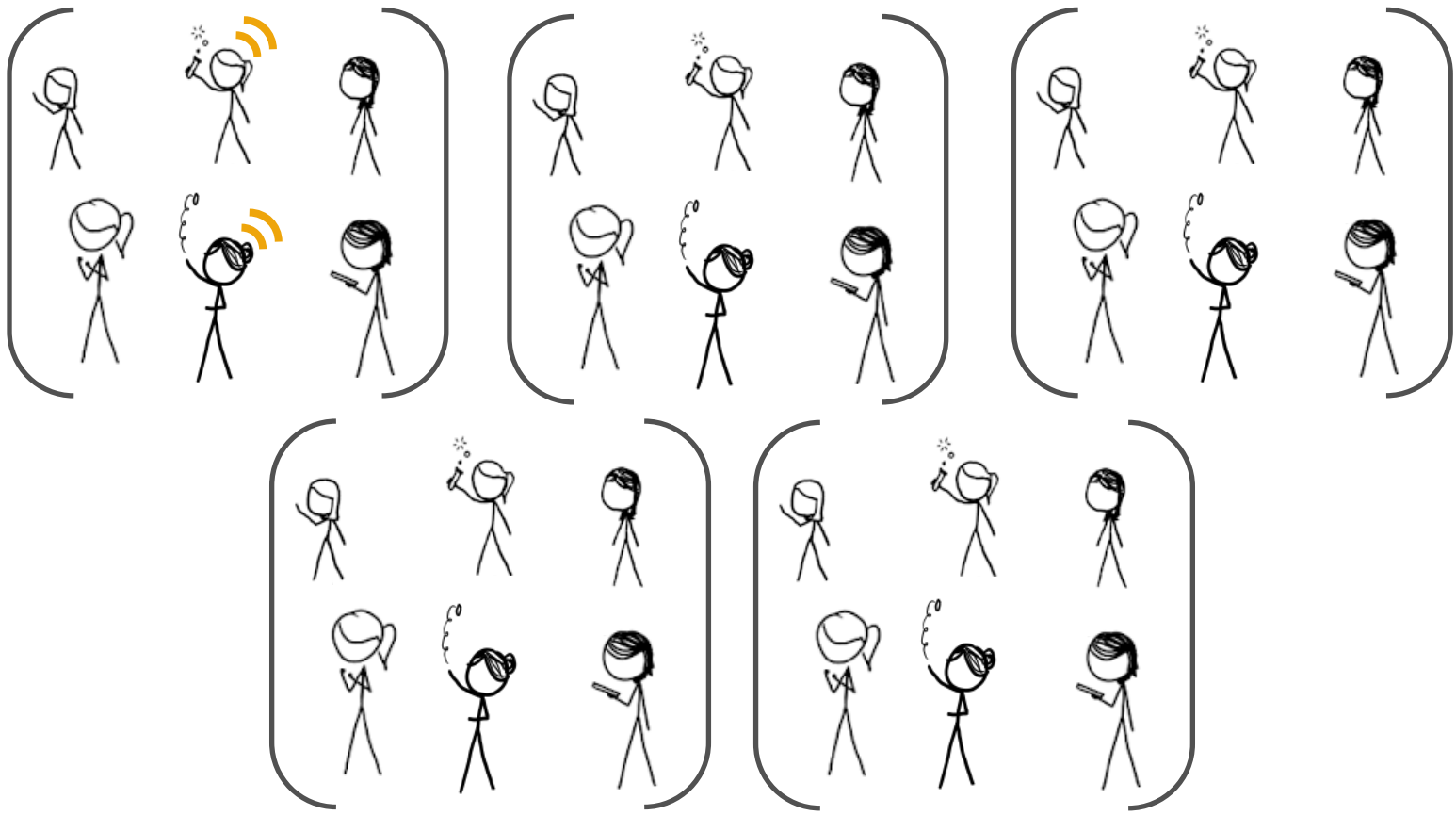


Who is Responsible for Creating Content?

- How many users are responsible for how many posts?

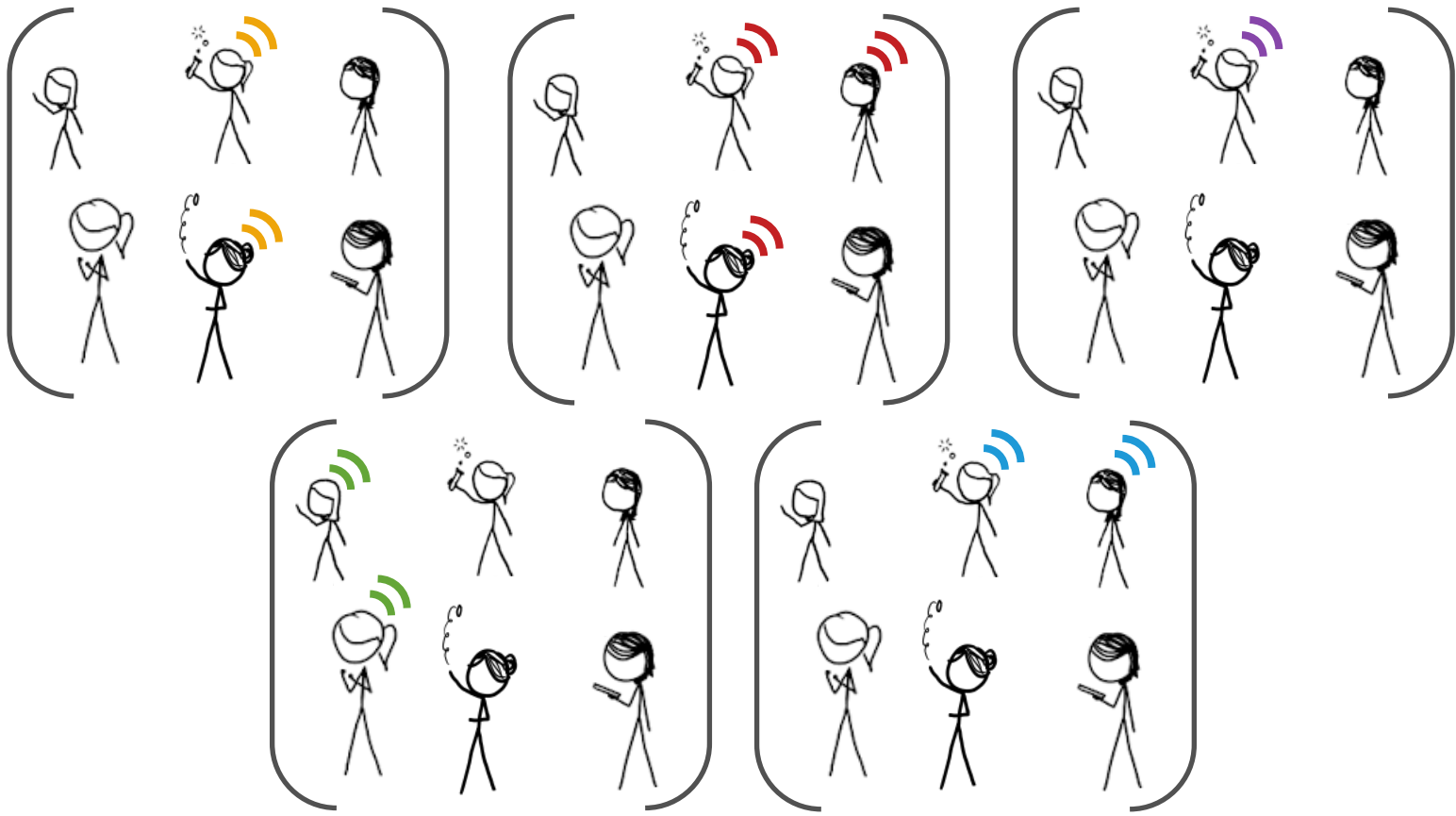
Who is Responsible for Creating Content?

- How many users are responsible for how many posts?

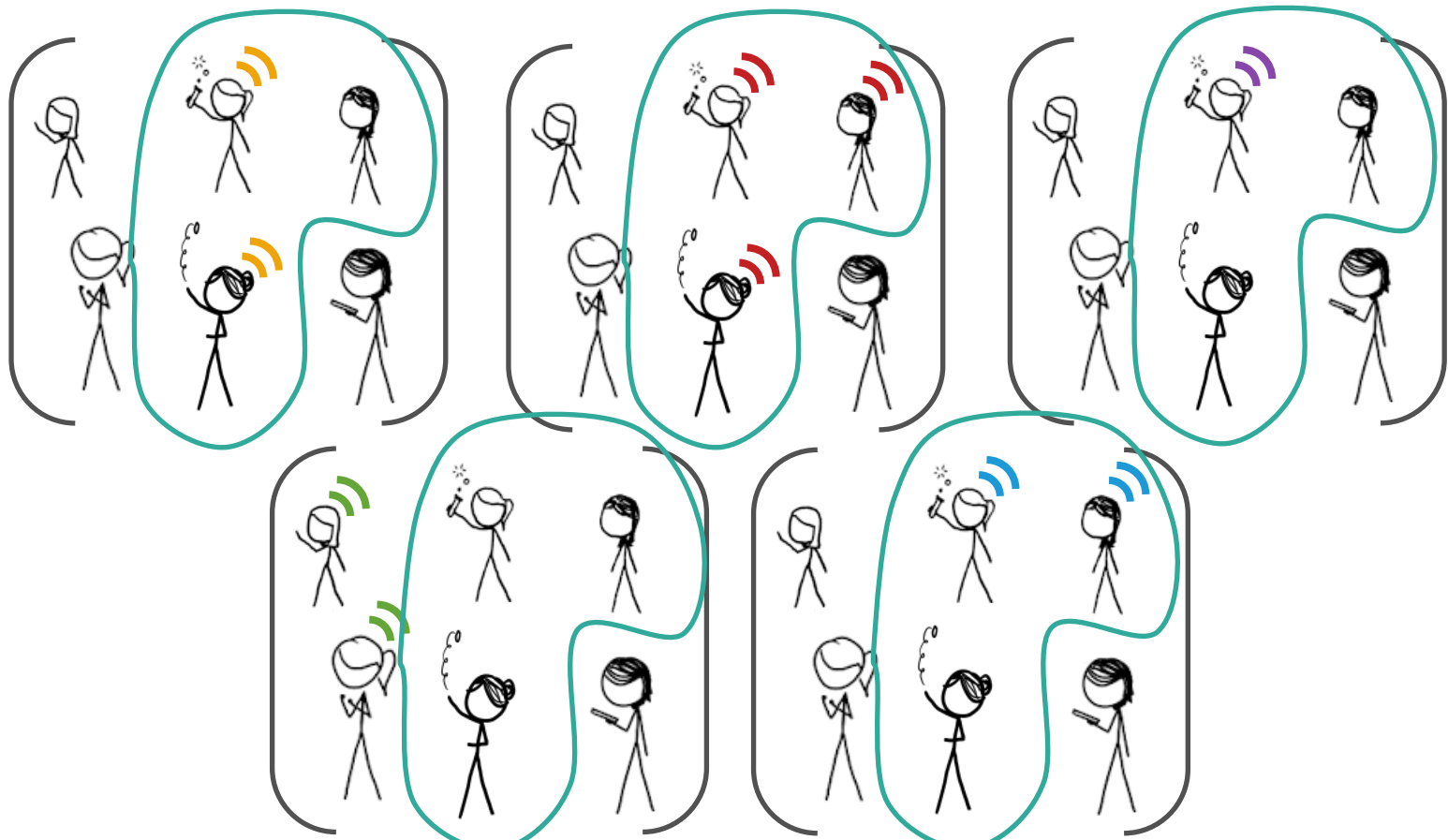


Who is Responsible for Creating Content?

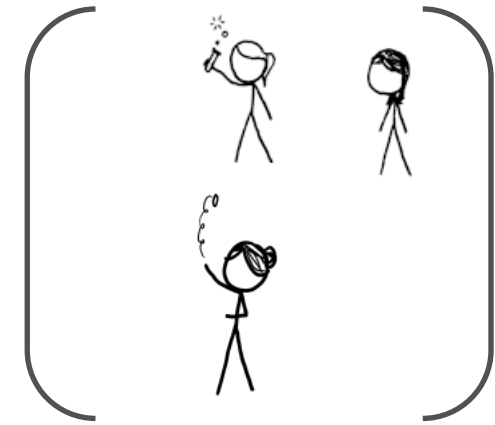
- How many users are responsible for how many posts?



Who is Responsible for Creating Content?



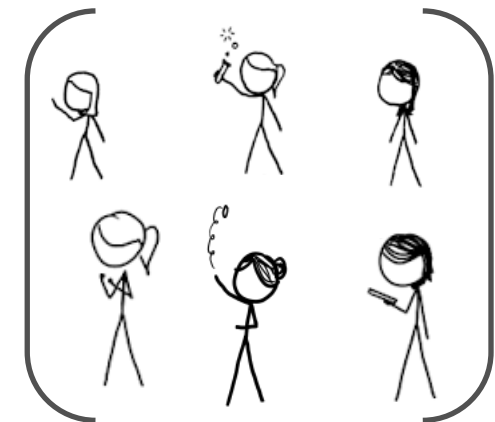
Who is Responsible for Creating Content?



= 50% of users
are responsible for

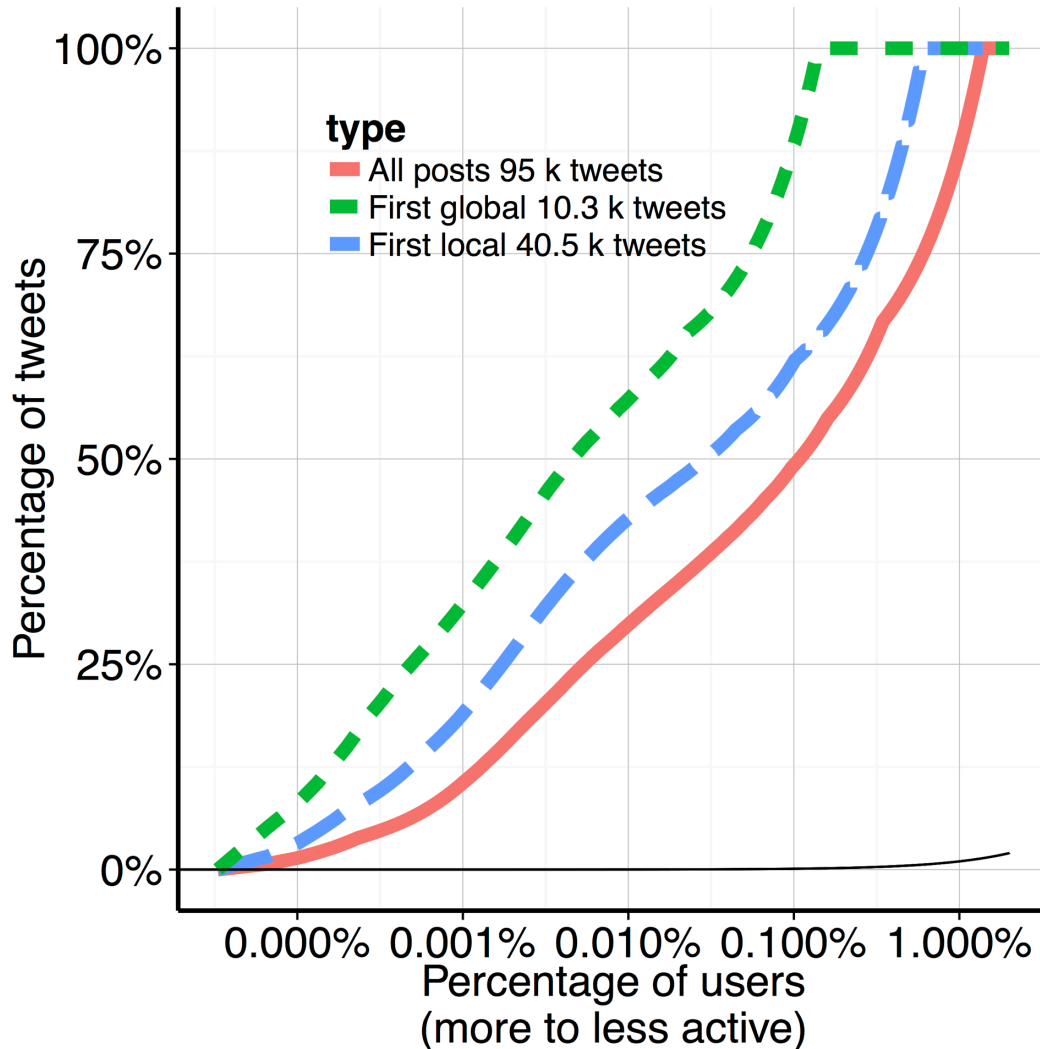


= 80% of posts

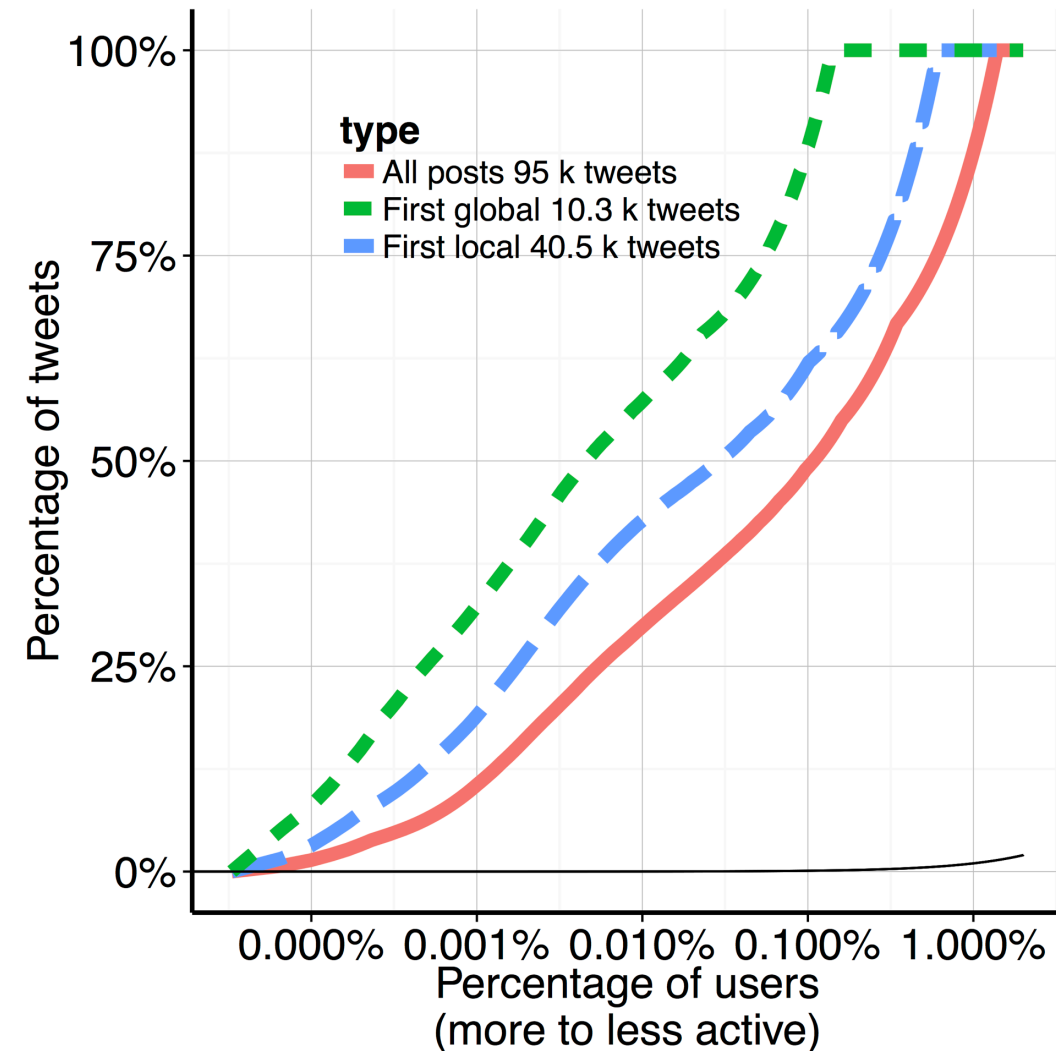


Who is Responsible for Creating Content?

Single domain
(cnn.com)



Who is Responsible for Creating Content?



Smaller fraction of users responsible for first tweets :

More original content is *more concentrated*

But what happens to other domains?

- Compare different types of domains
 - nytimes.com
 - Daily, Shorter lifespan

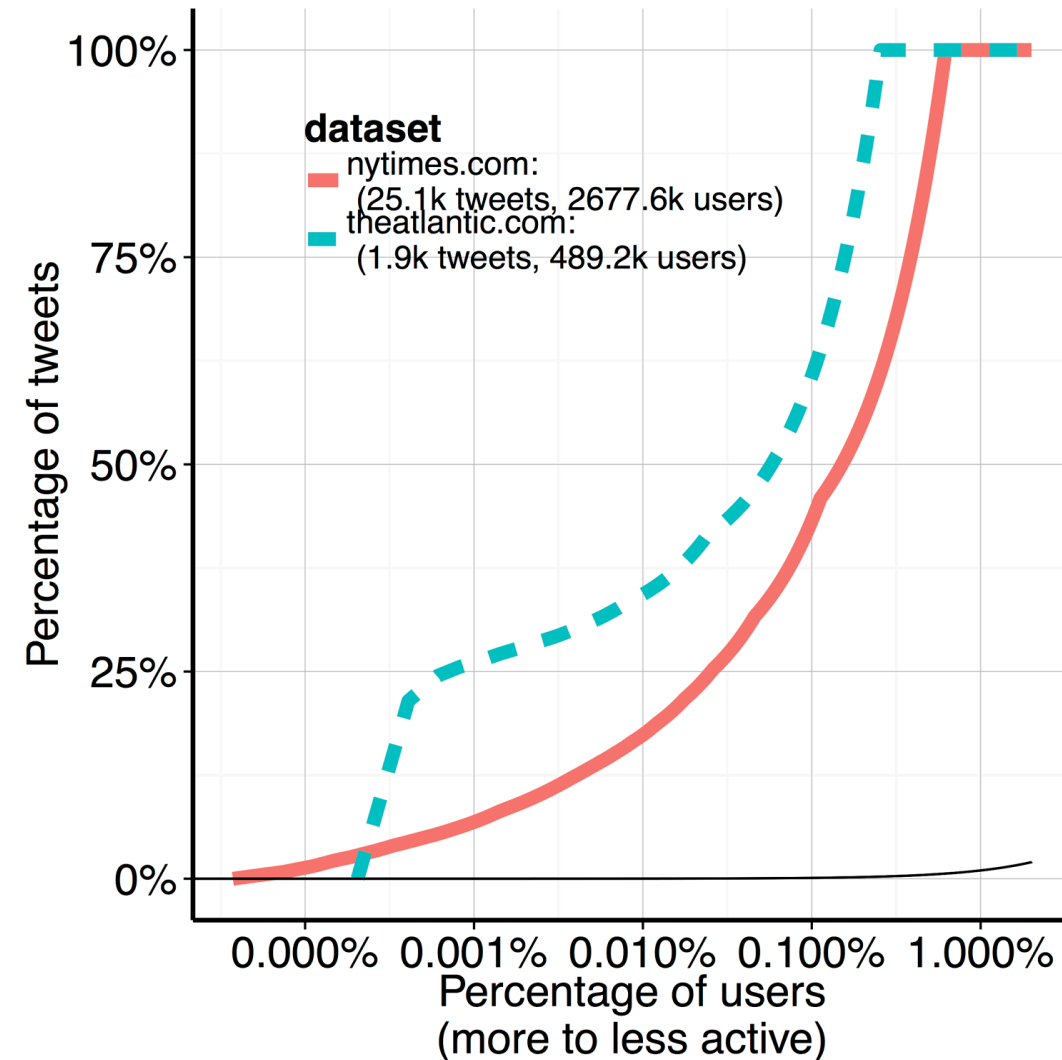


– theatlantic.com

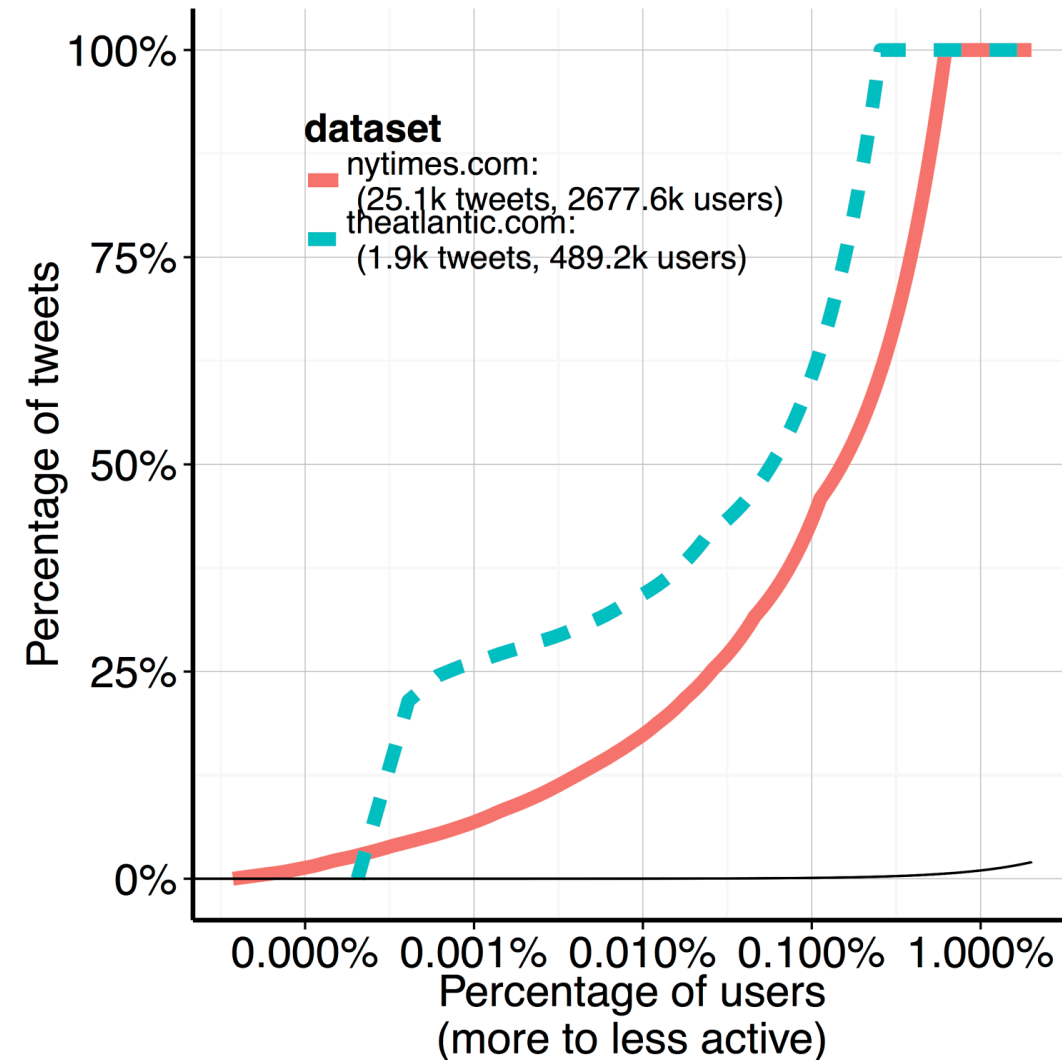
- Monthly, Longer lifespan



But what happens to other domains?

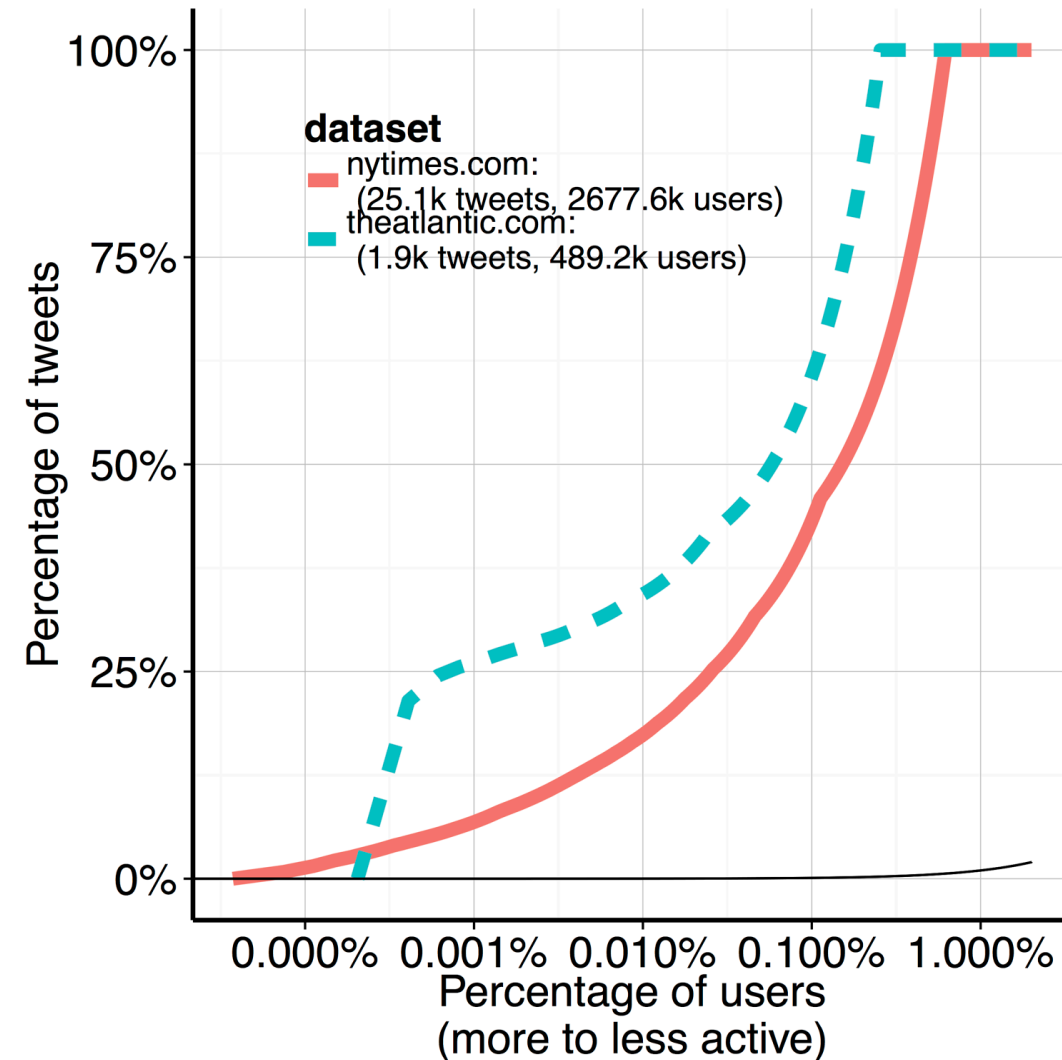


But what happens to other domains?



Shorter lifespan is
less concentrated

But what happens to other domains?



Shorter lifespan is
less concentrated

Longer lifespan is
more concentrated

Outline

- Datasets
- Who are the Content Creators?
- Relationship of Content Lifespan and Concentration
- Model of Perishable Public Goods
- Equilibria and Specialization
- Conclusion

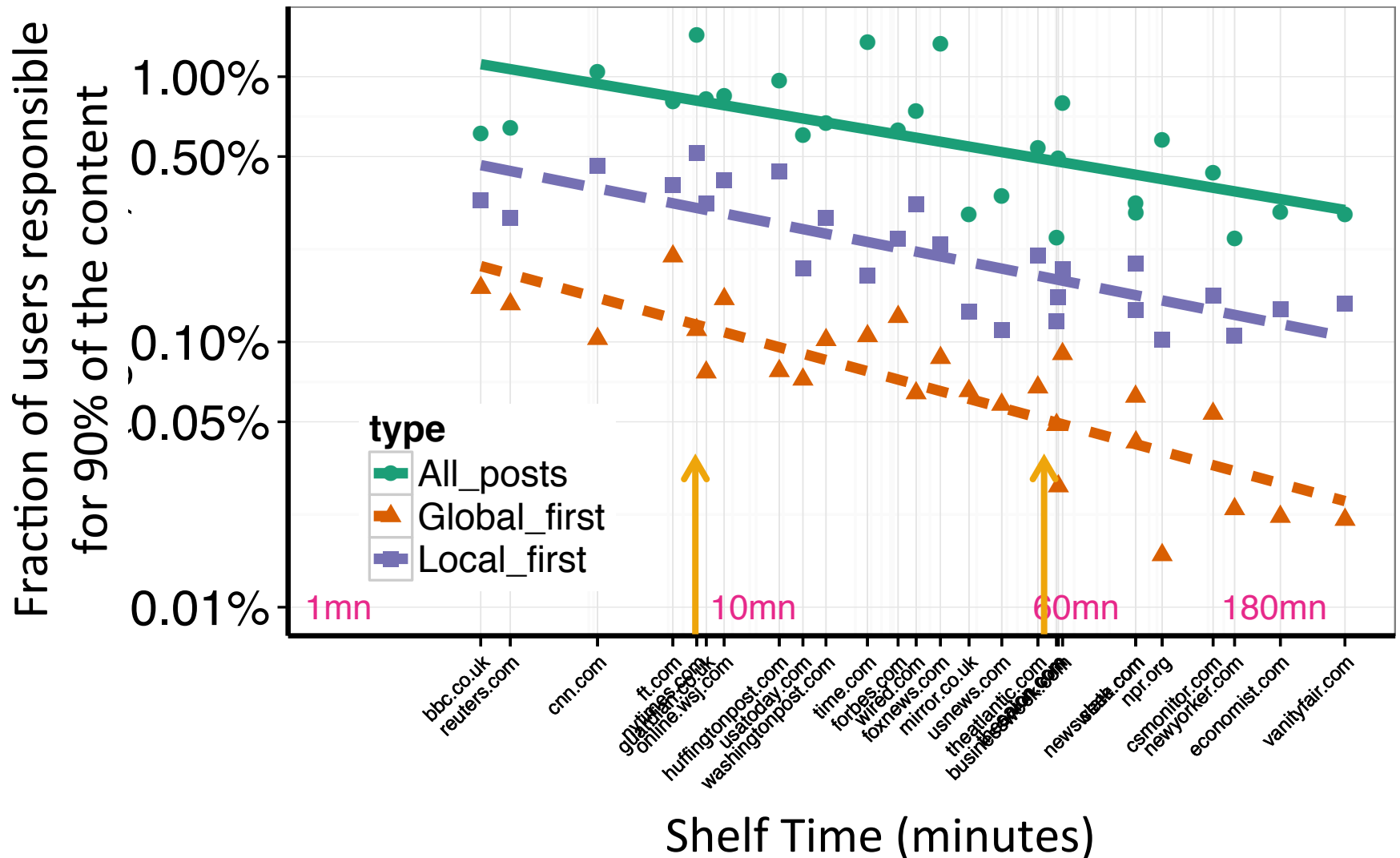
What's the relationship between lifespan and concentration?

- Measure of lifespan : *Shelf life*
 - Expected attention for an article before it is replaced
 - Multiple ways to measure
 - based on volume of tweets
 - based on duration

What's the relationship between lifespan and concentration?

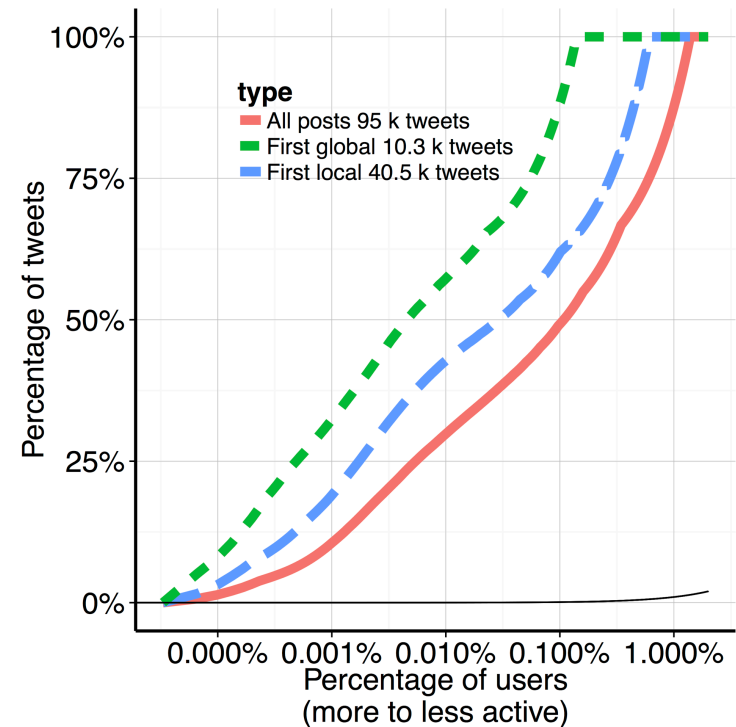
- Measure of lifespan : *Shelf life*
 - Expected attention for an article before it is replaced
 - Multiple ways to measure
 - based on volume of tweets
 - based on duration
 - Here we use the one based on volume

Concentration of sharing for different media sources



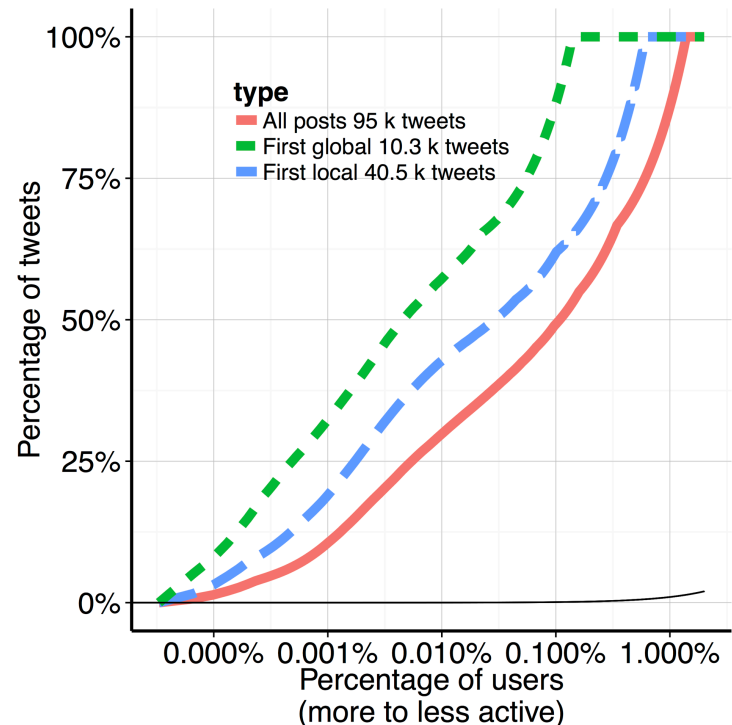
Observations

- Specialization exists



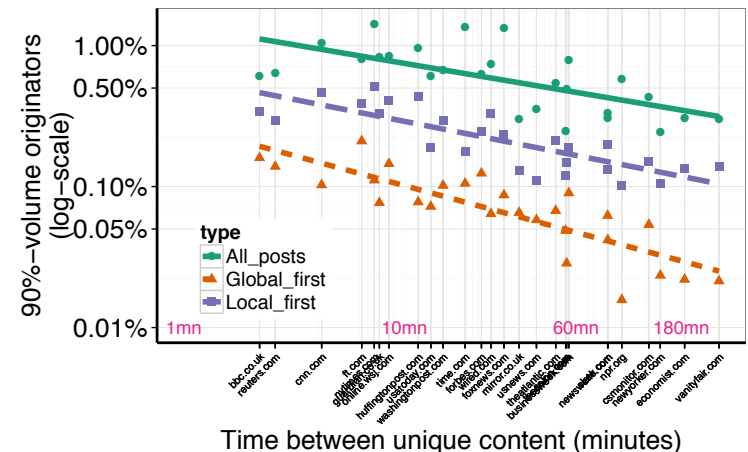
Observations

- Specialization exists
- Understanding who contributes is not trivial
 - Eg: Original content doesn't come from the highest degree nodes



Observations

- Specialization exists
- Understanding who contributes is not trivial
 - Eg: Original content doesn't come from the highest degree nodes
- Time is an important factor
 - Short lived content reduces specialization



Observations

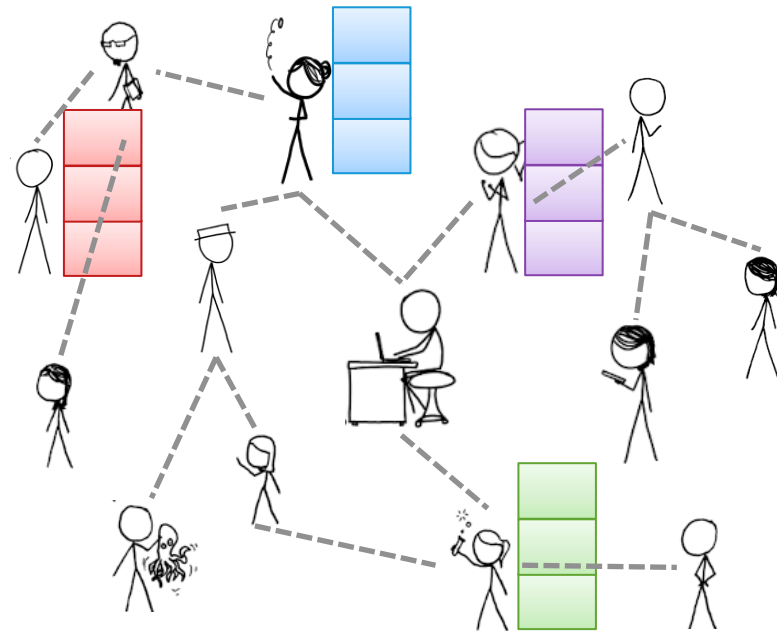
- Specialization exists
- Understanding who contributes is not trivial
 - Eg: Original content doesn't come from the highest degree nodes
- Time is an important factor
 - Short lived content reduces specialization
- What are the conditions under which specialization occurs?
 - Formally?
 - What dynamics causes this effect?

Outline

- Datasets
- Who are the Content Creators?
- Relationship of Content Lifespan and Concentration
- **Model of Perishable Public Goods**
- Equilibria and Specialization
- Conclusion

Perishable Public Goods Model

- Properties of the model
 - Specialization exists
 - Understanding who contributes is not trivial
 - Eg: Original content doesn't come from the highest degree nodes
 - Time is an important factor
 - Short lived content reduces specialization



Finding Information



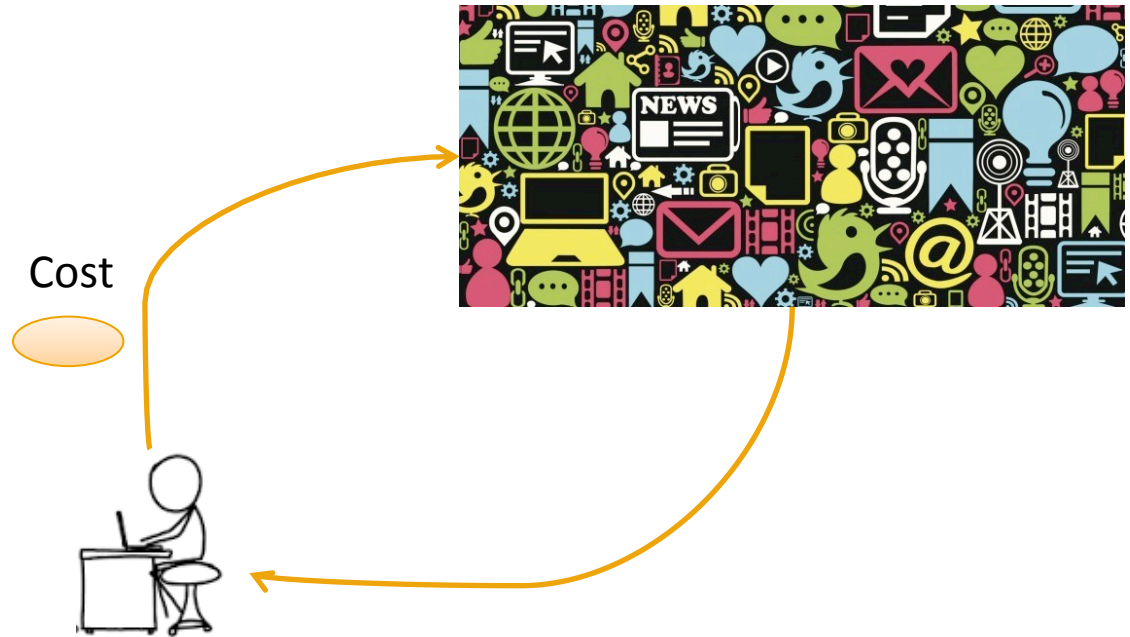
Finding Information



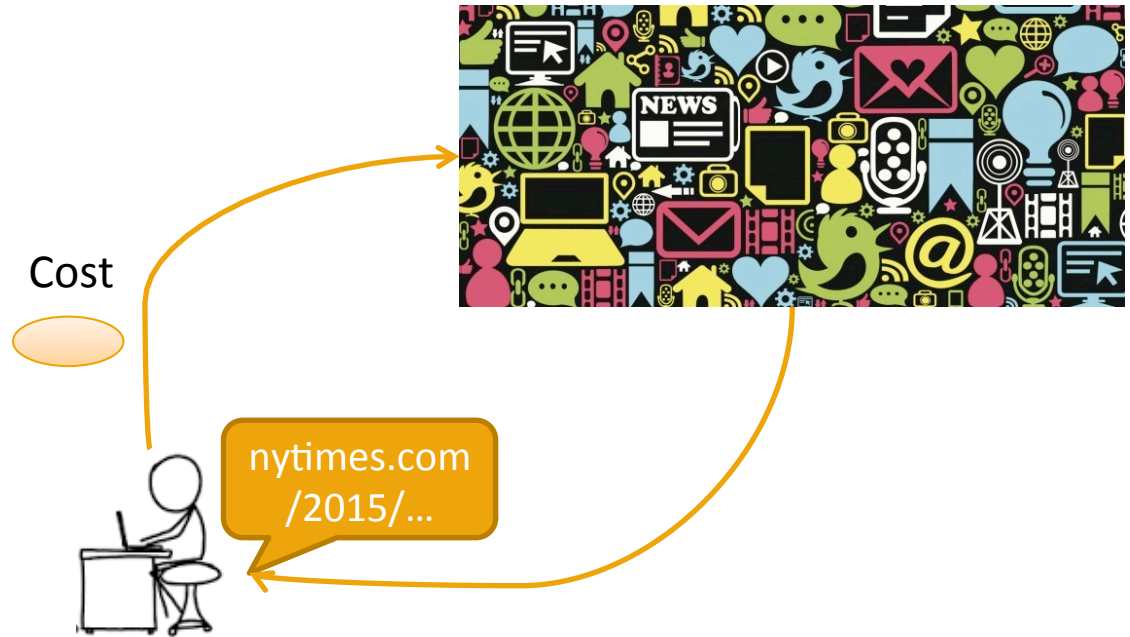
Finding Information



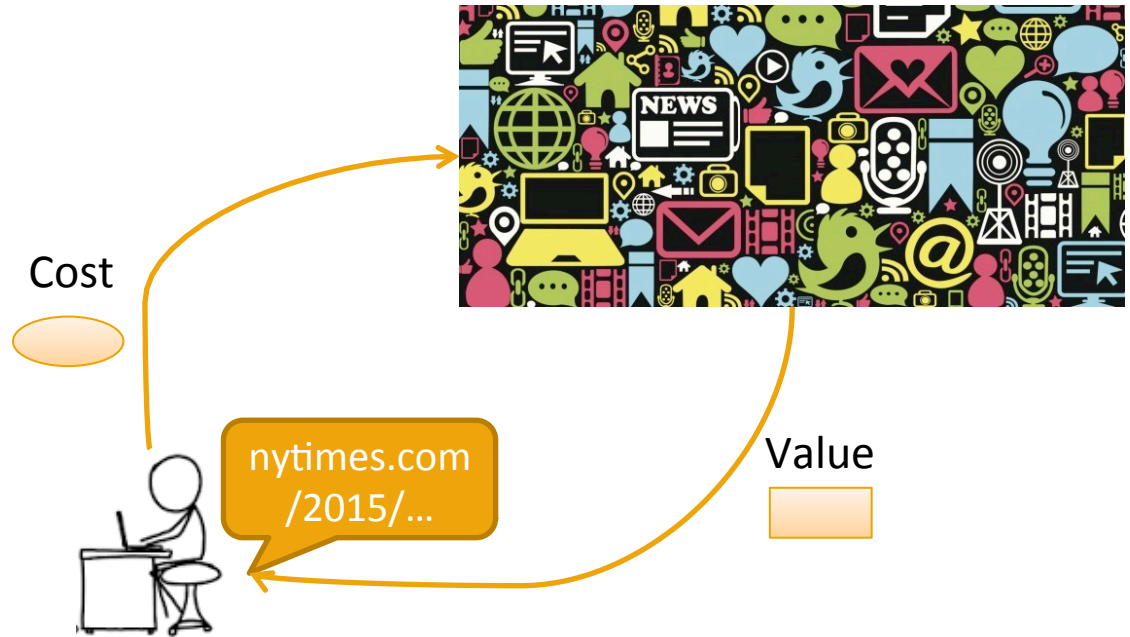
Finding Information



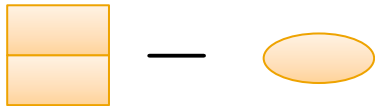
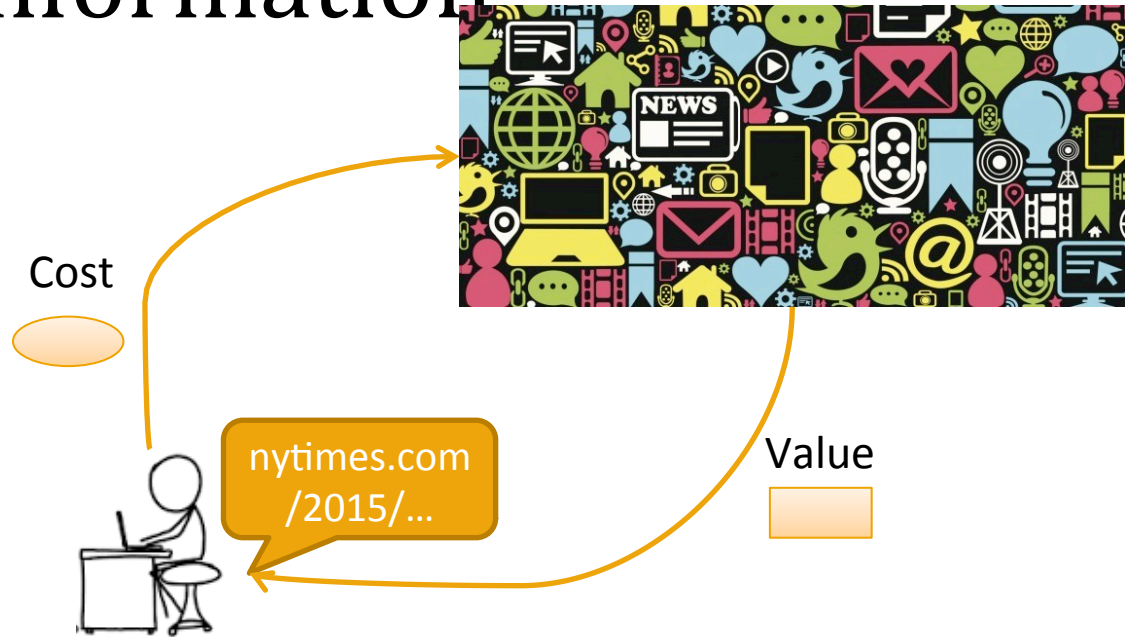
Finding Information



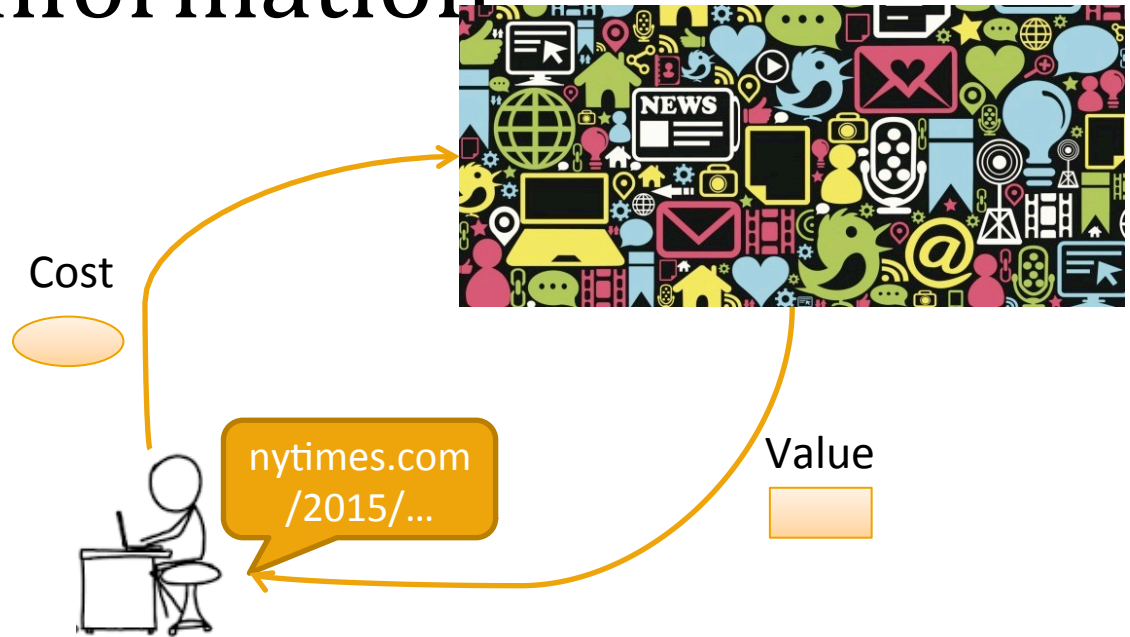
Finding Information



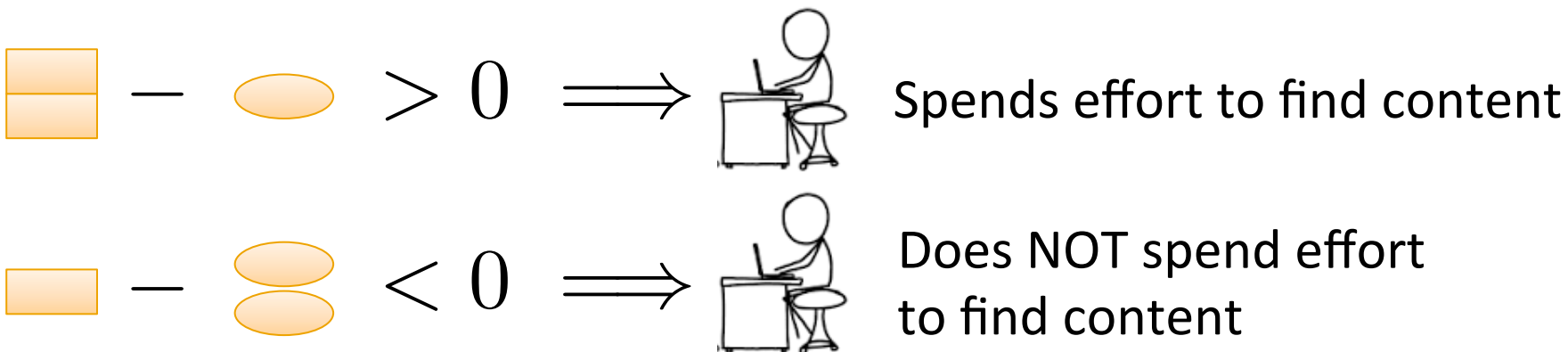
Choosing Investment to Find Information



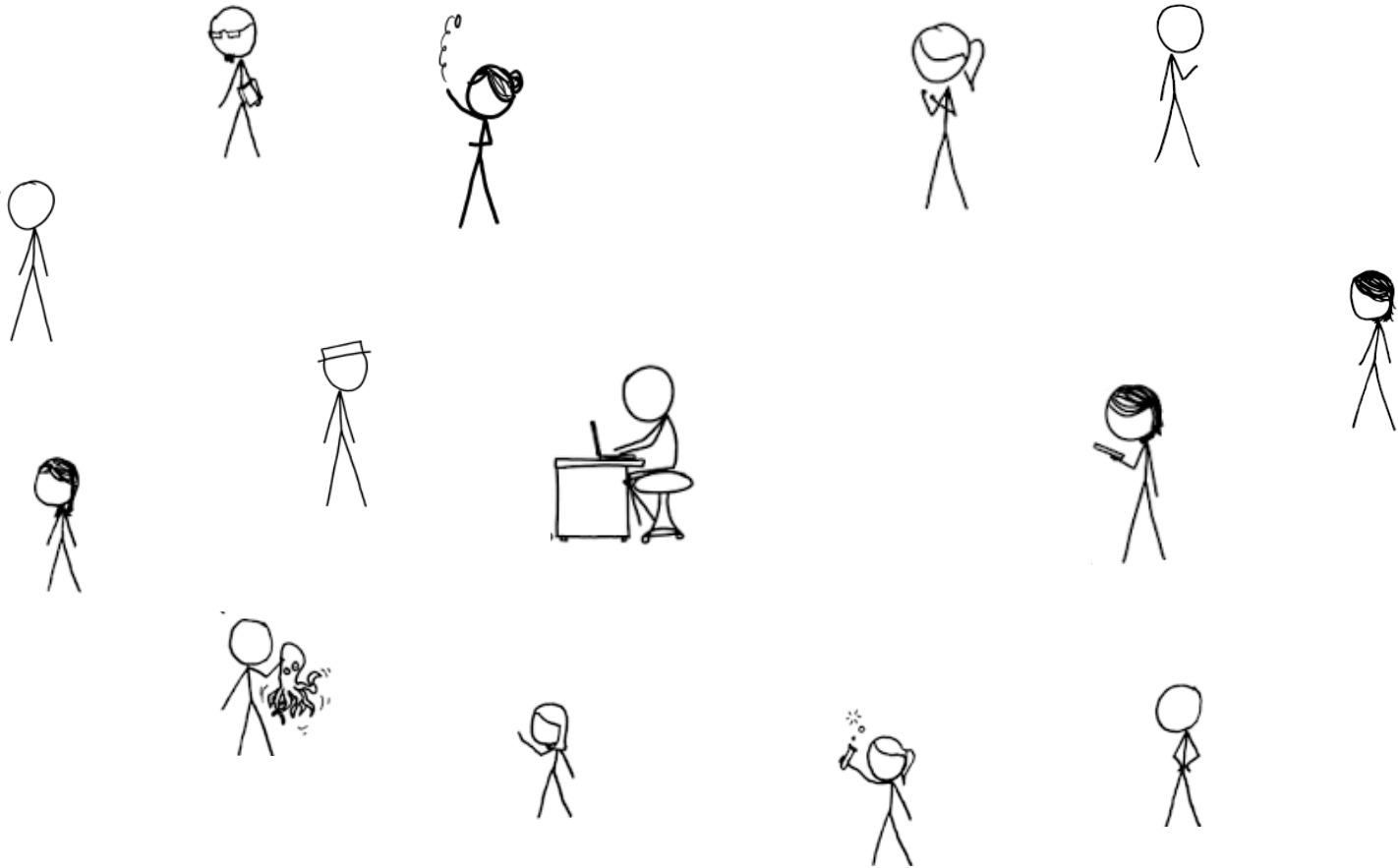
Choosing Investment to Find Information



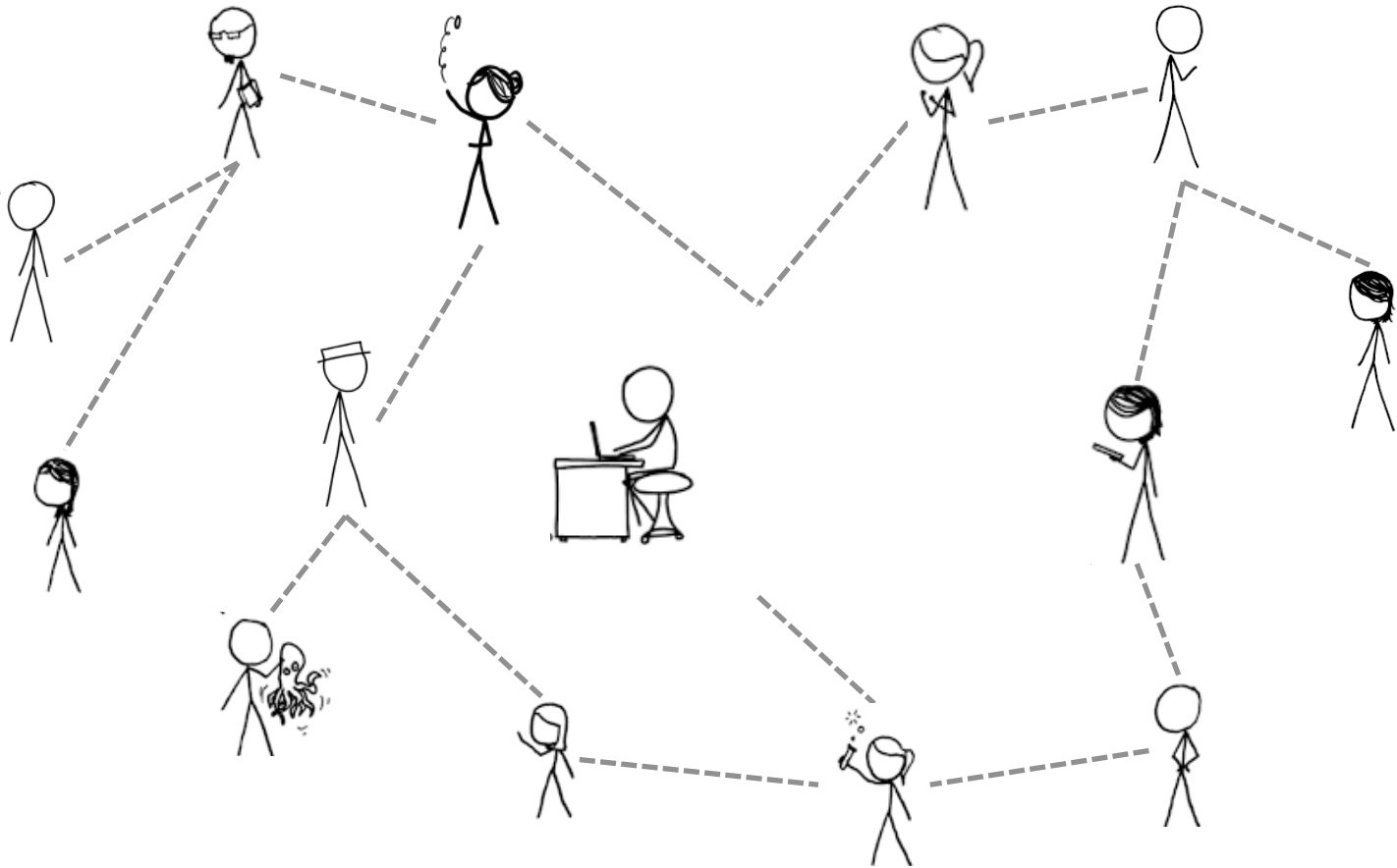
$$\text{Value} - \text{Cost} > 0 \implies \text{Stick Figure at Laptop} \quad \text{Spends effort to find content}$$



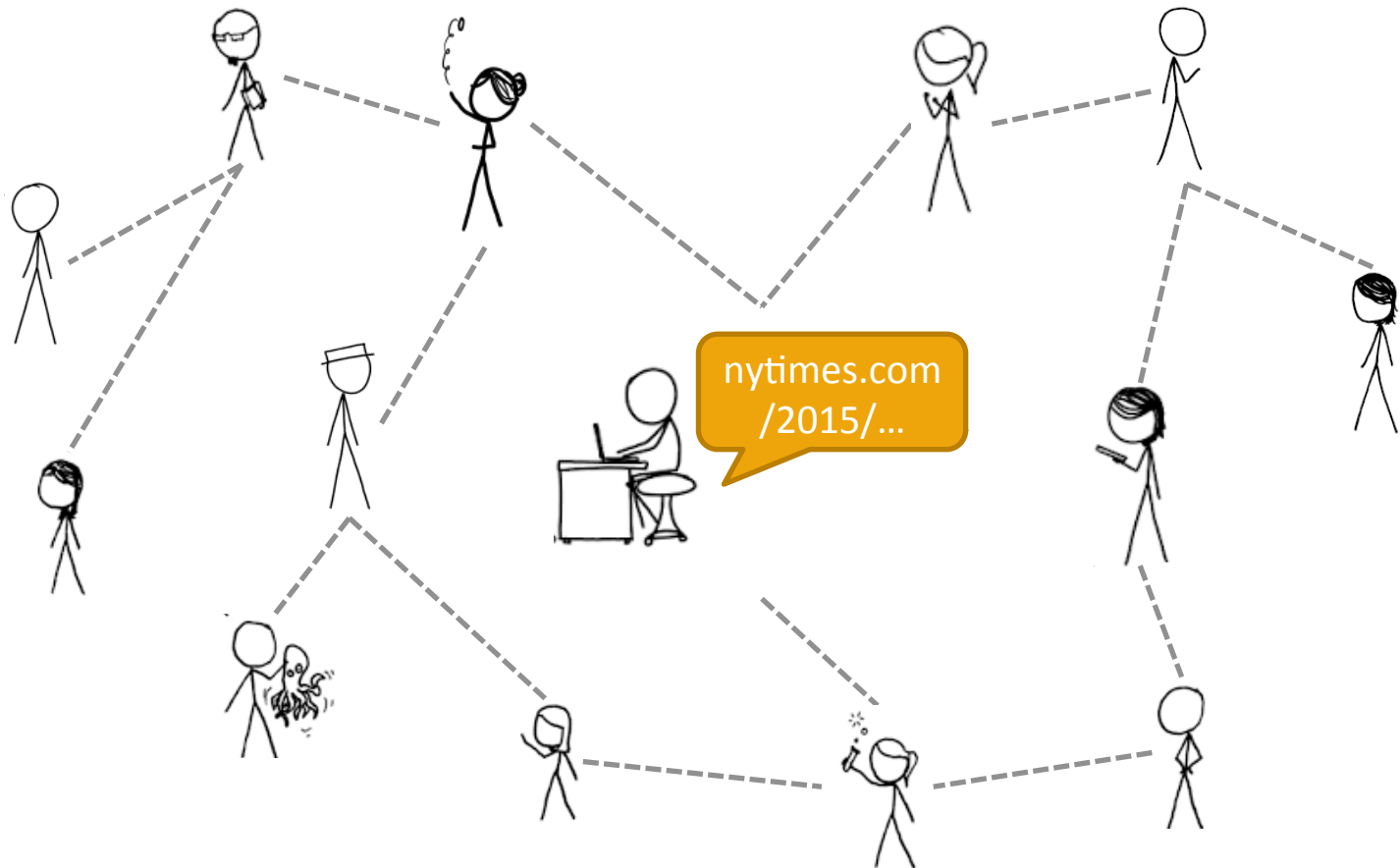
Public Goods Model



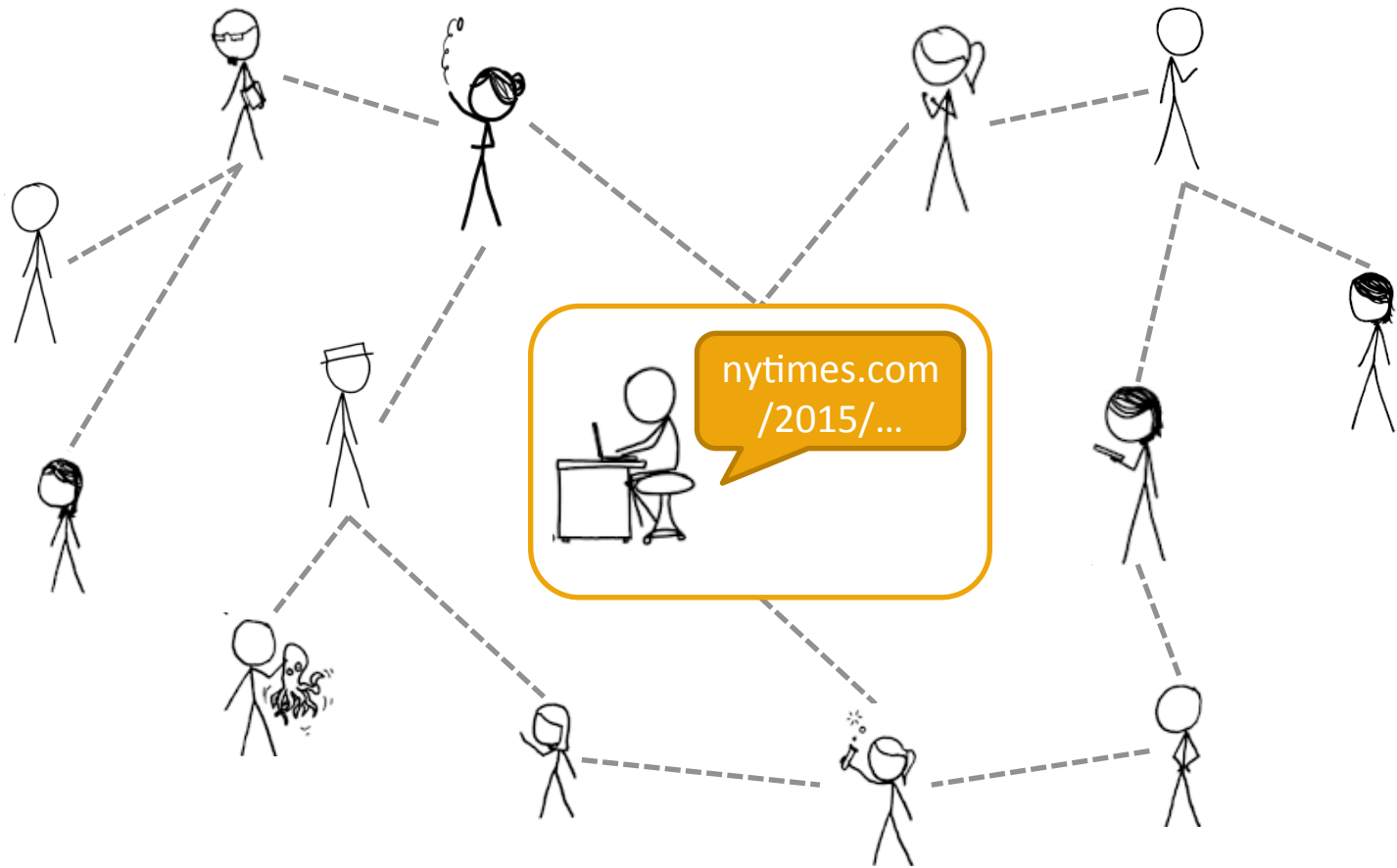
Perishable Public Goods Model



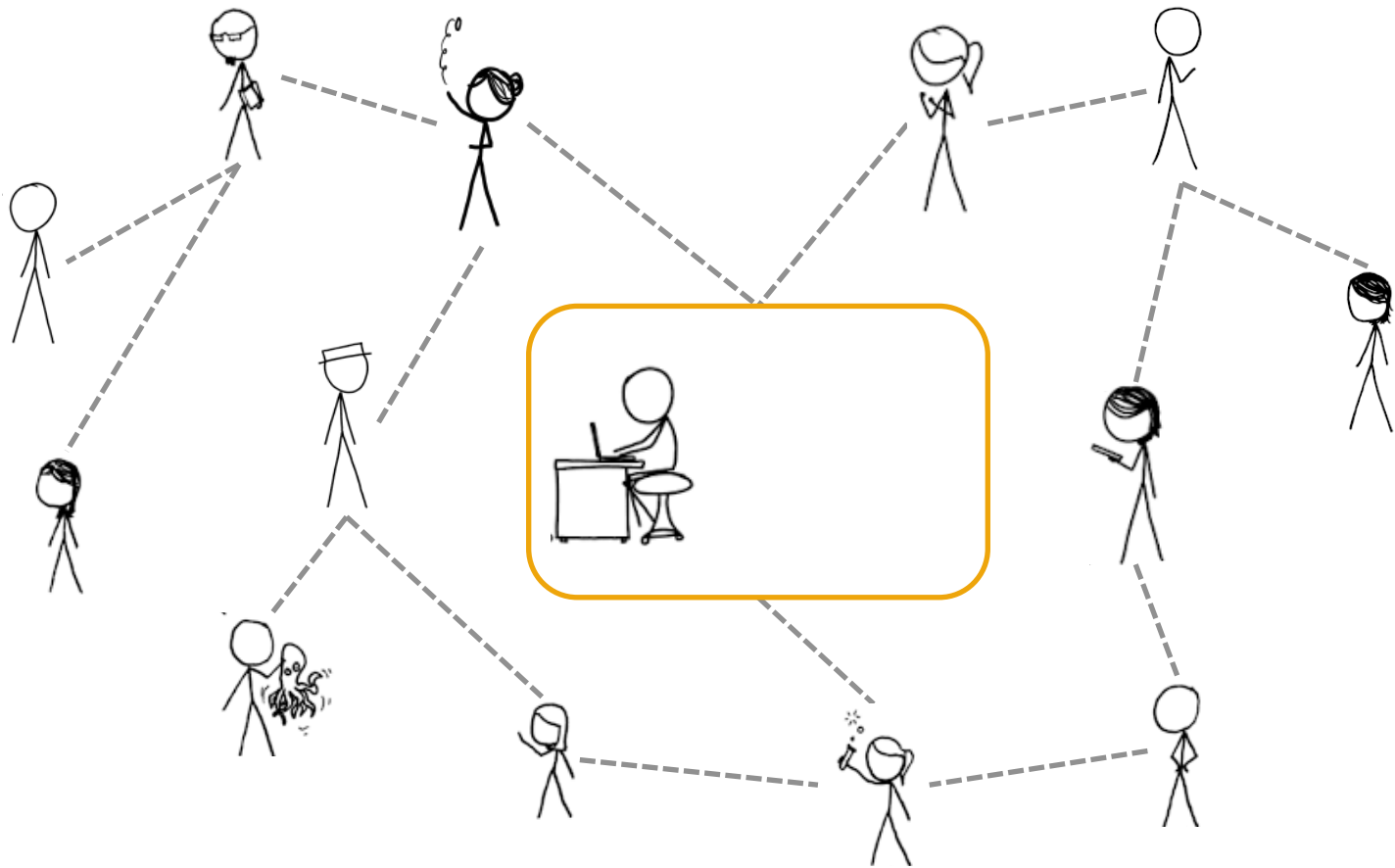
Perishable Public Goods Model



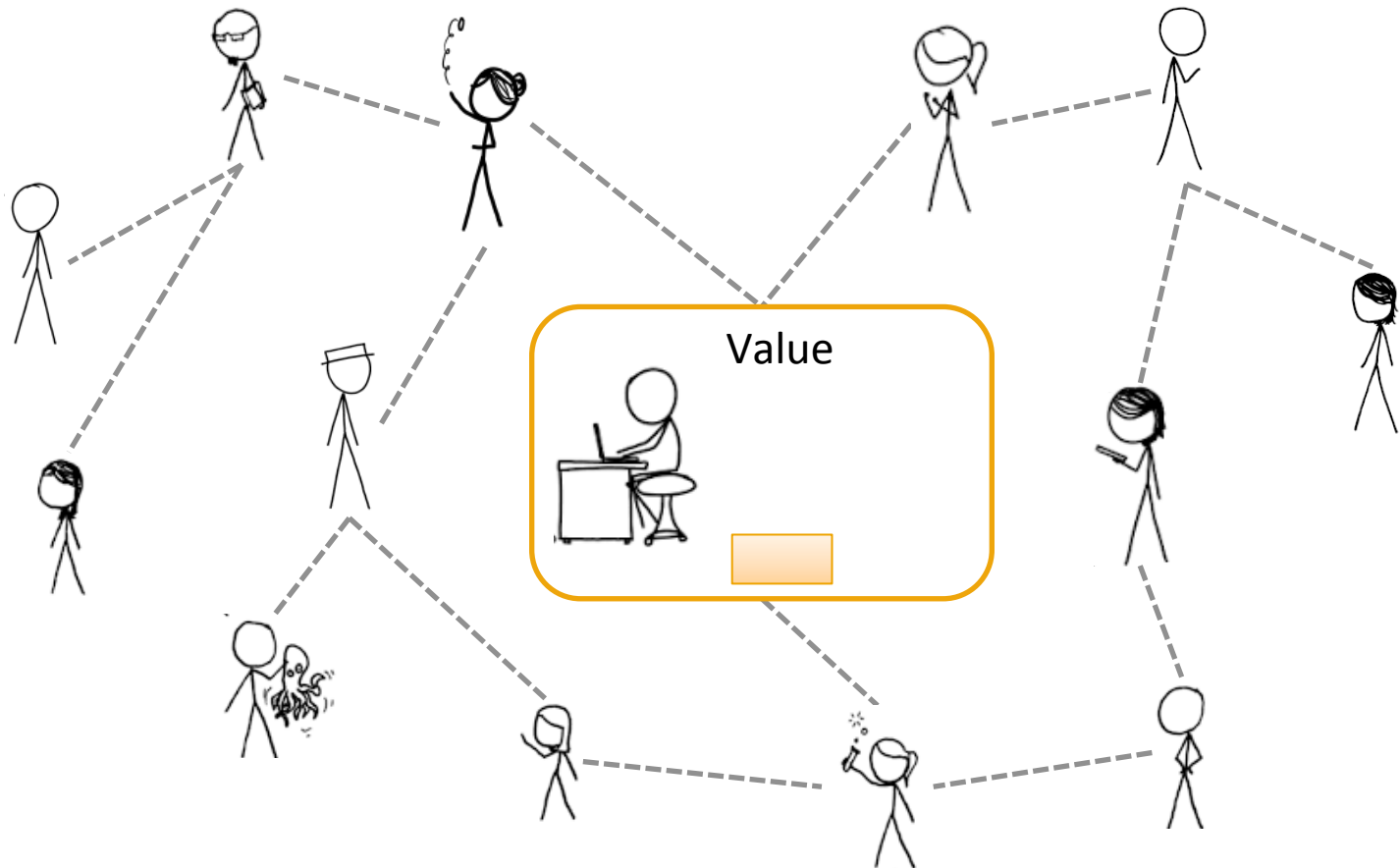
Perishable Public Goods Model



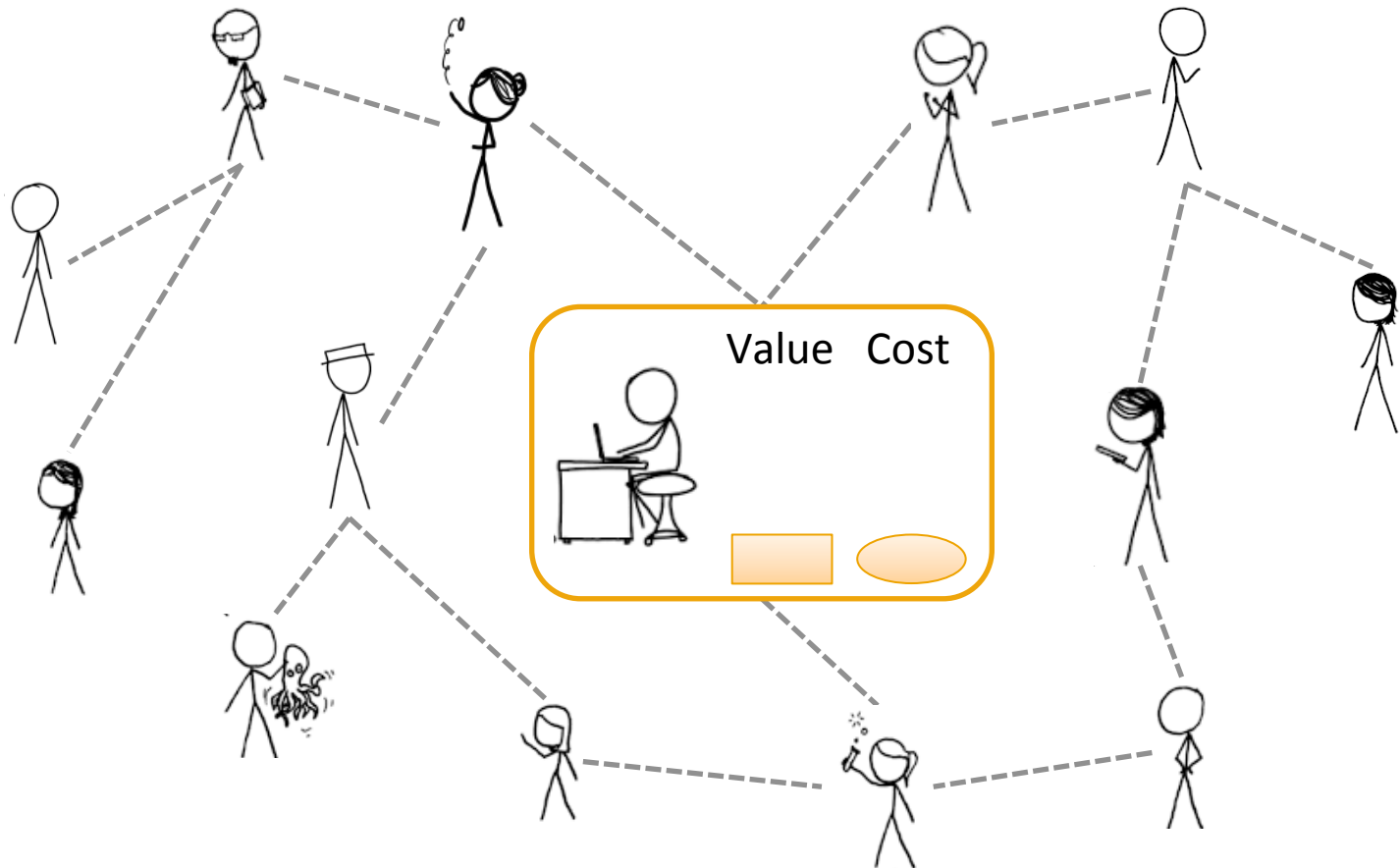
Perishable Public Goods Model



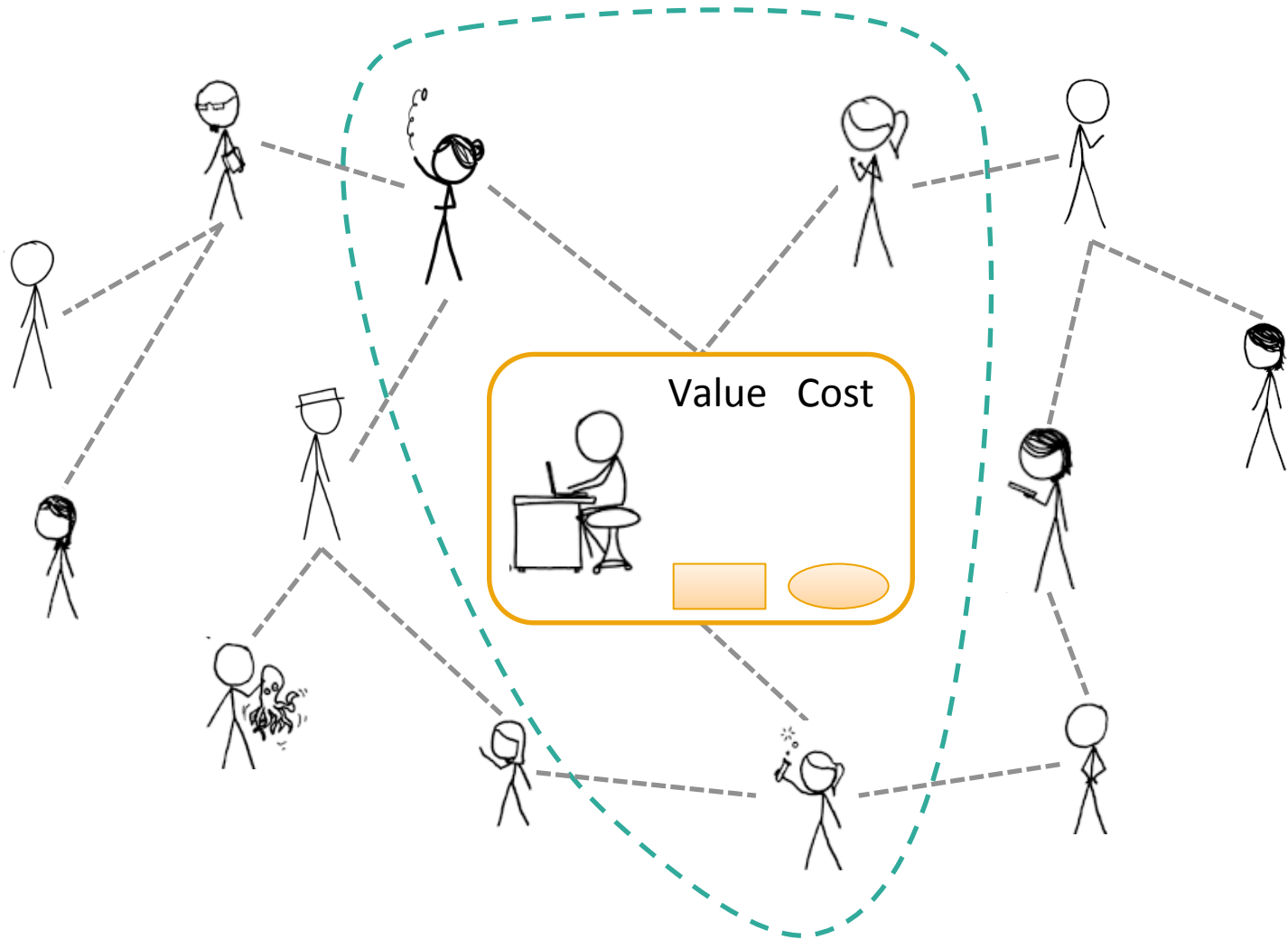
Perishable Public Goods Model



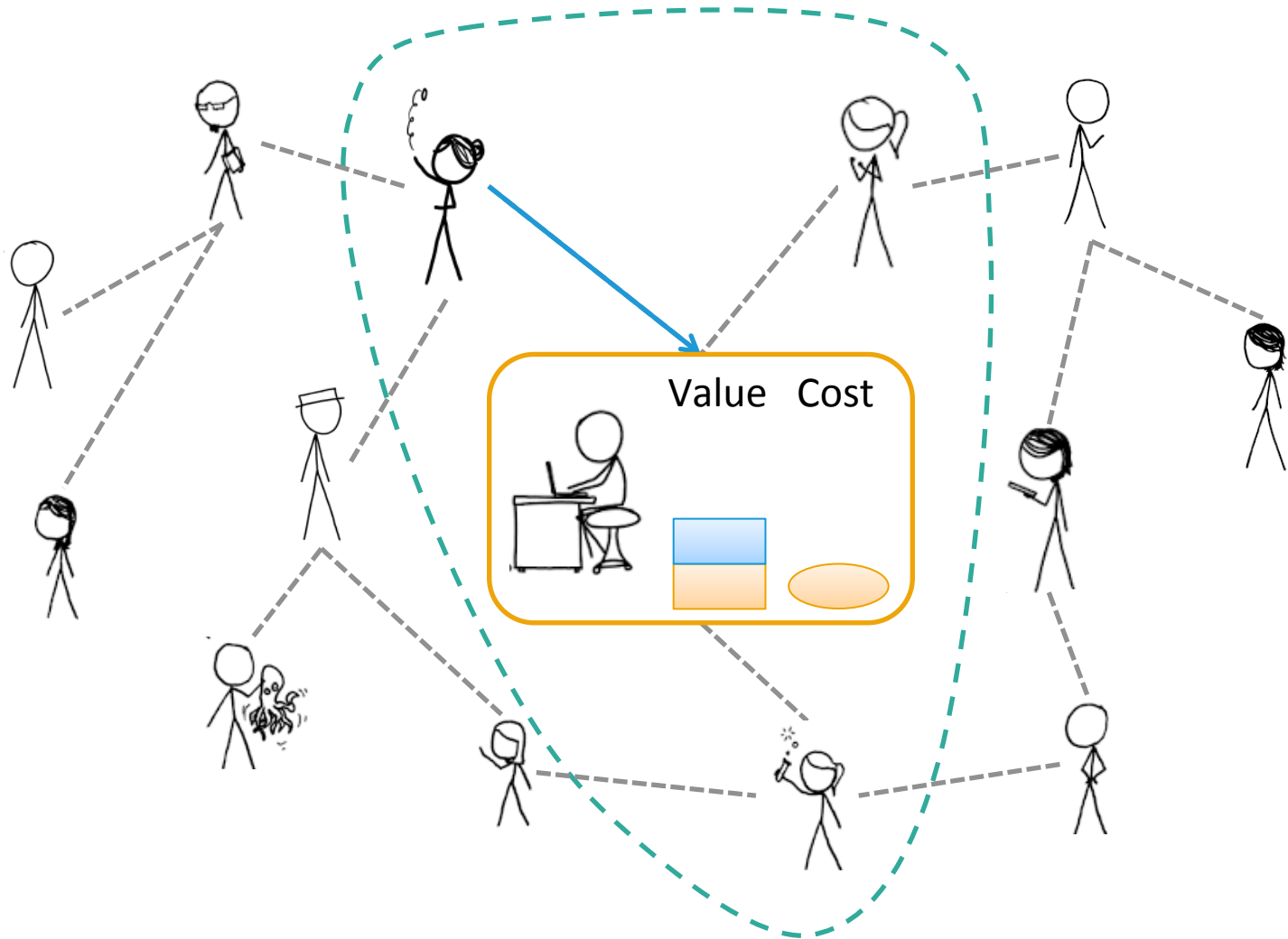
Perishable Public Goods Model



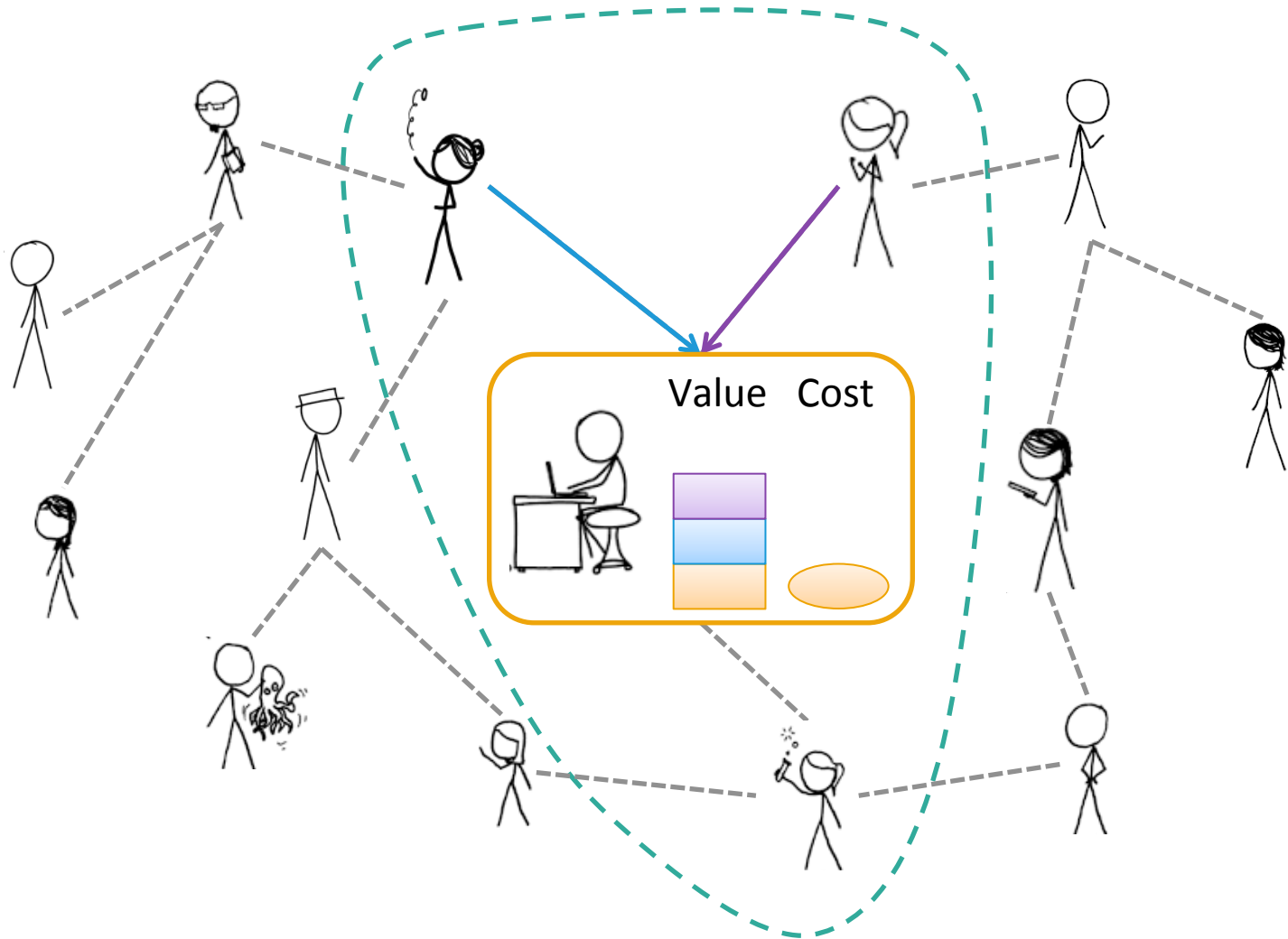
Perishable Public Goods Model



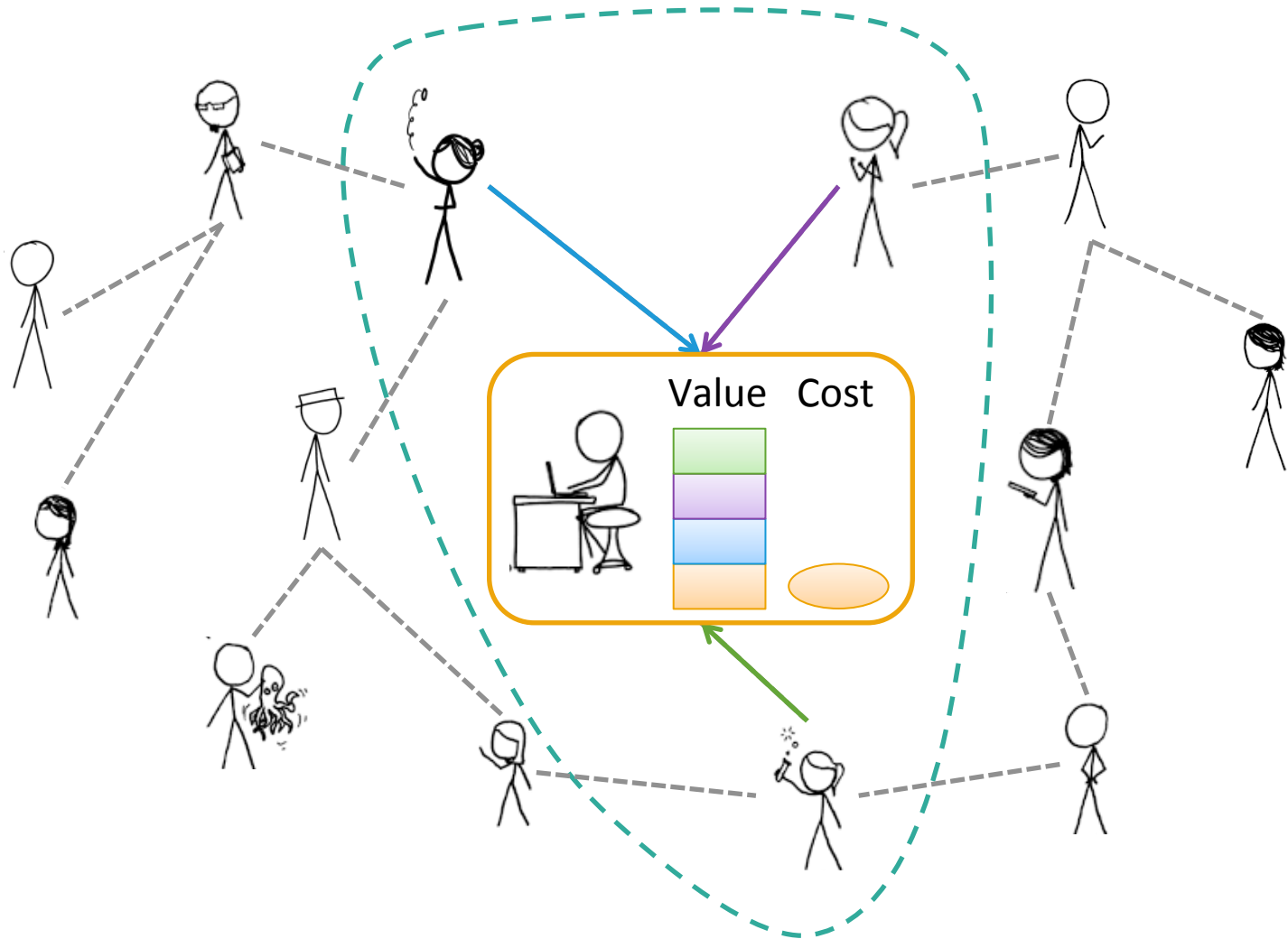
Perishable Public Goods Model



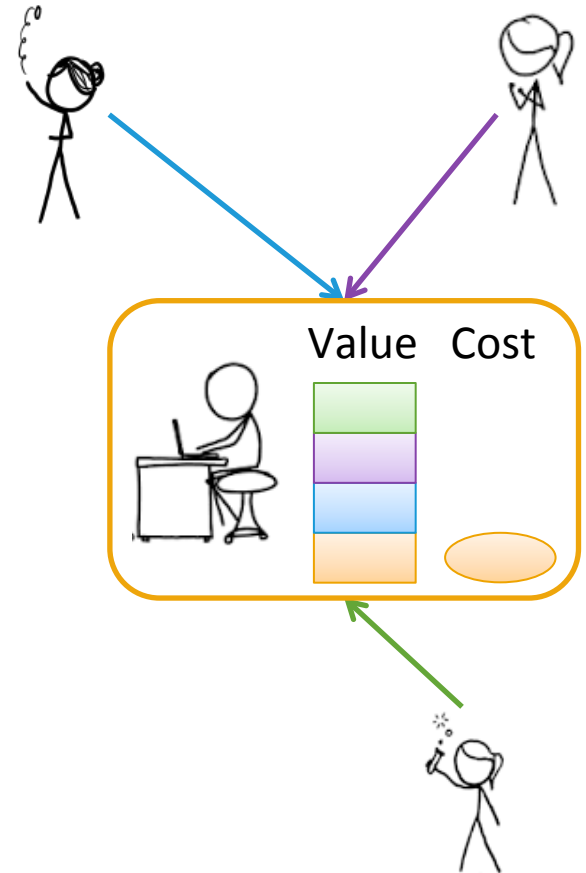
Perishable Public Goods Model



Perishable Public Goods Model




Perishable Public Goods Model

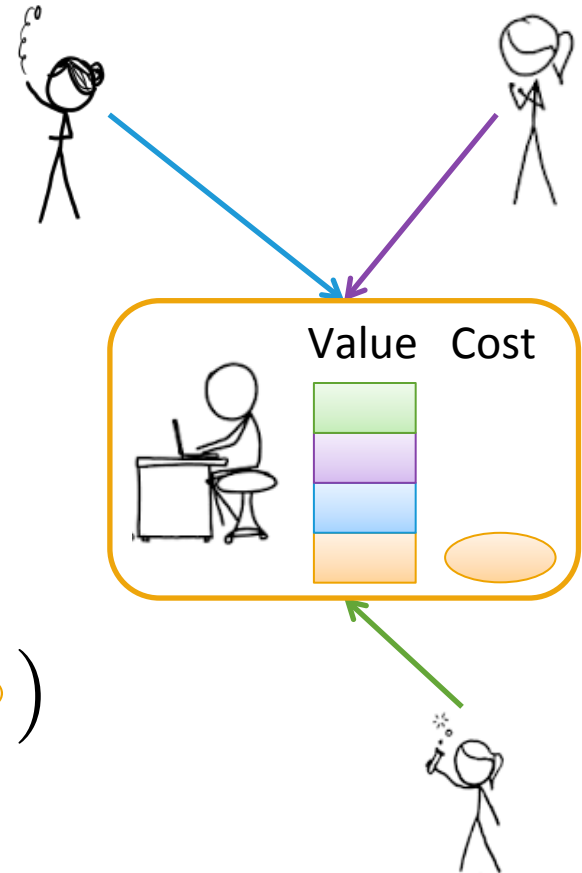


Perishable Public Goods Model

$$y_i =$$



$$y_{-i} =$$


$$U(y_i, y_{-i}) = f\left(\begin{array}{c} \text{green} \\ \text{purple} \\ \text{blue} \\ \text{orange} \end{array}\right) - c(\text{orange oval})$$

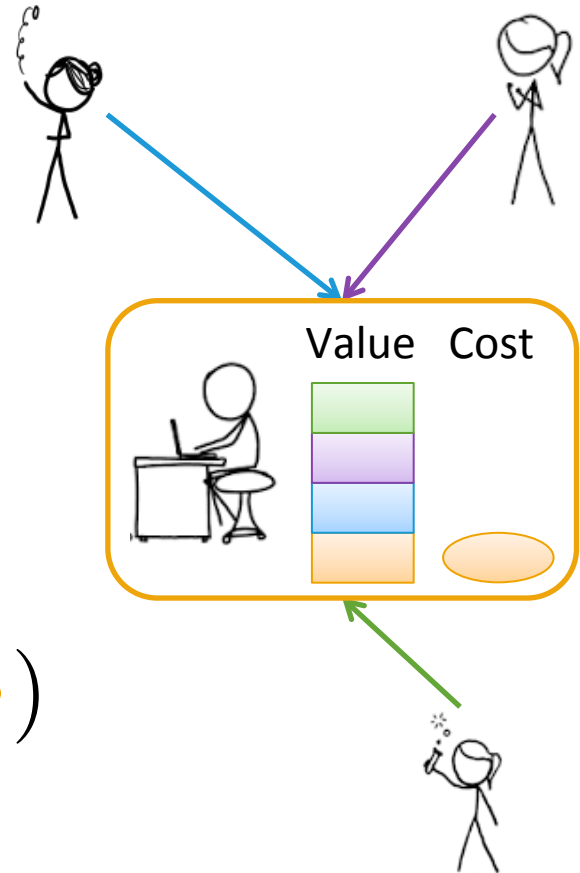


Public Goods Model

$$y_i =$$


$$y_{-i} =$$


$$U(y_i, y_{-i}) = f\left(\begin{array}{c} \text{green} \\ \text{purple} \\ \text{blue} \\ \text{orange} \end{array}\right) - c(\text{orange oval})$$

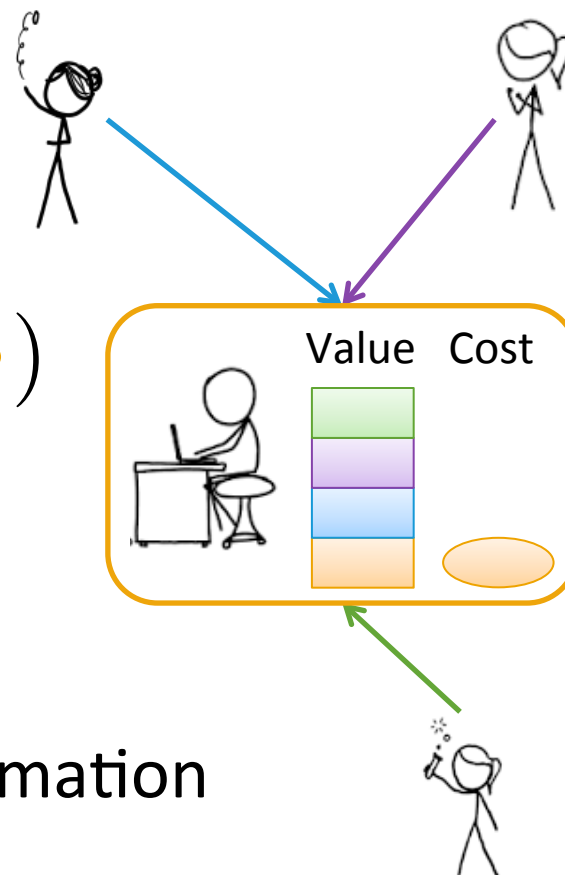


Public Goods Model

$$U(y_i, y_{-i}) = f\left(\begin{array}{c} \text{green box} \\ \text{purple box} \\ \text{blue box} \\ \text{orange box} \end{array}\right) - c(\text{orange oval})$$

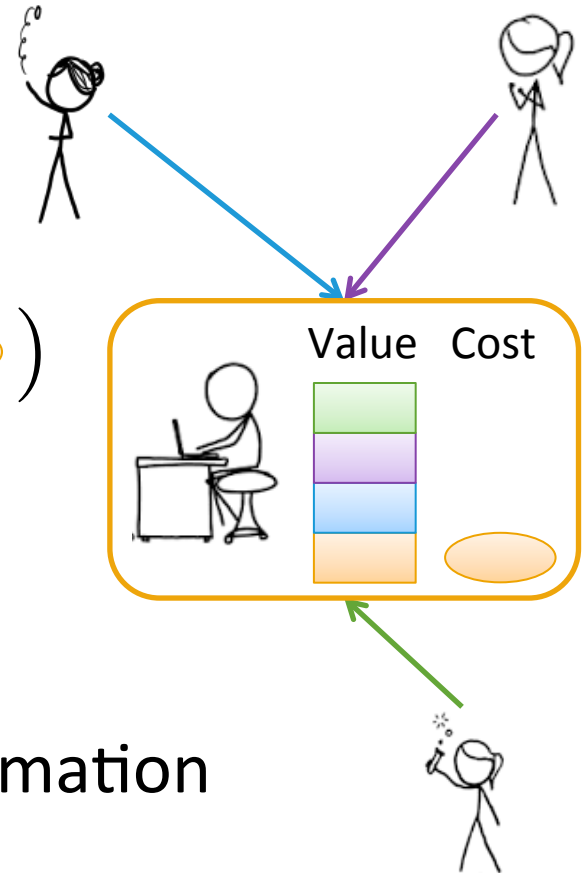
Assume :

- an article has a shelf life of τ
- y_i is the rate of discovery of information



Perishable Public Goods Model

$$U(y_i, y_{-i}) = f\left(\begin{array}{c} \text{green} \\ \text{purple} \\ \text{blue} \\ \text{orange} \end{array}\right) - c(\text{orange oval})$$



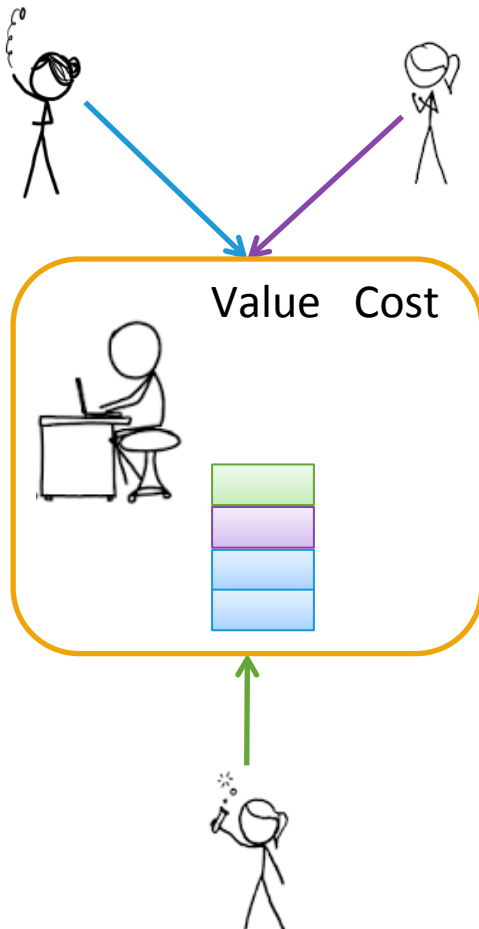
Assume :

- an article has a shelf life of τ
- y_i is the rate of discovery of information

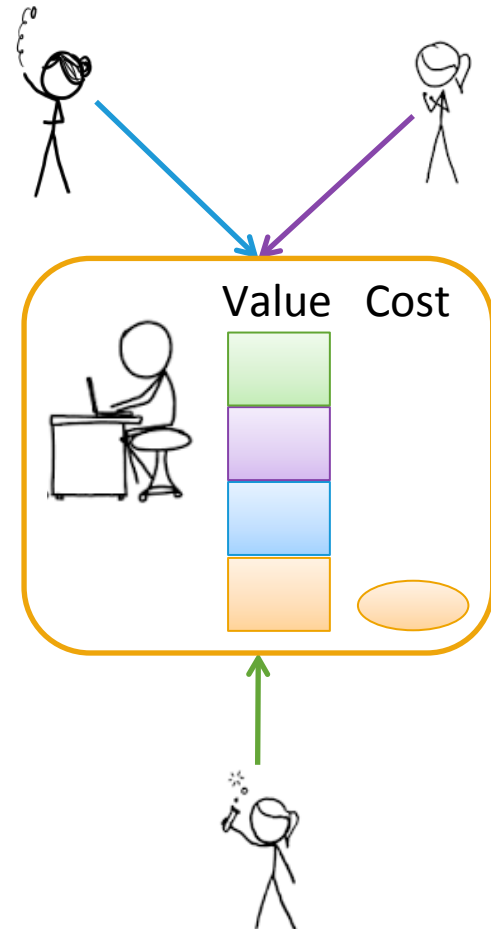
$$U(y_i, y_{-i}) = (1 - e^{\tau(y_i + y_{-i})}) - c(y_i)$$

Strategies for a user

Free-riding



Equitable distribution

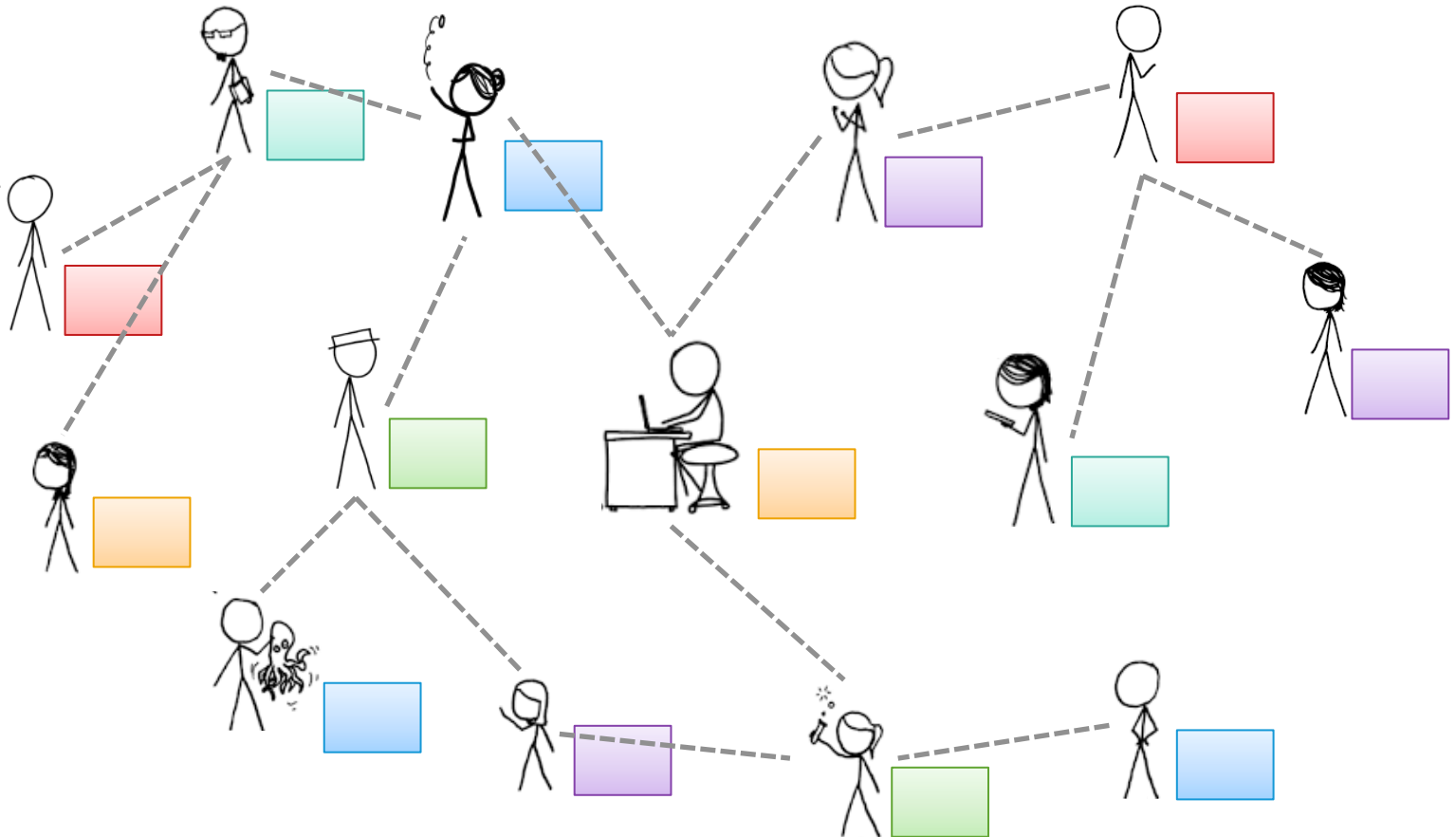


Outline

- Datasets
- Who are the Content Creators?
- Relationship of Content Lifespan and Concentration
- Model of Perishable Public Goods
- Equilibria and Specialization
- Conclusion

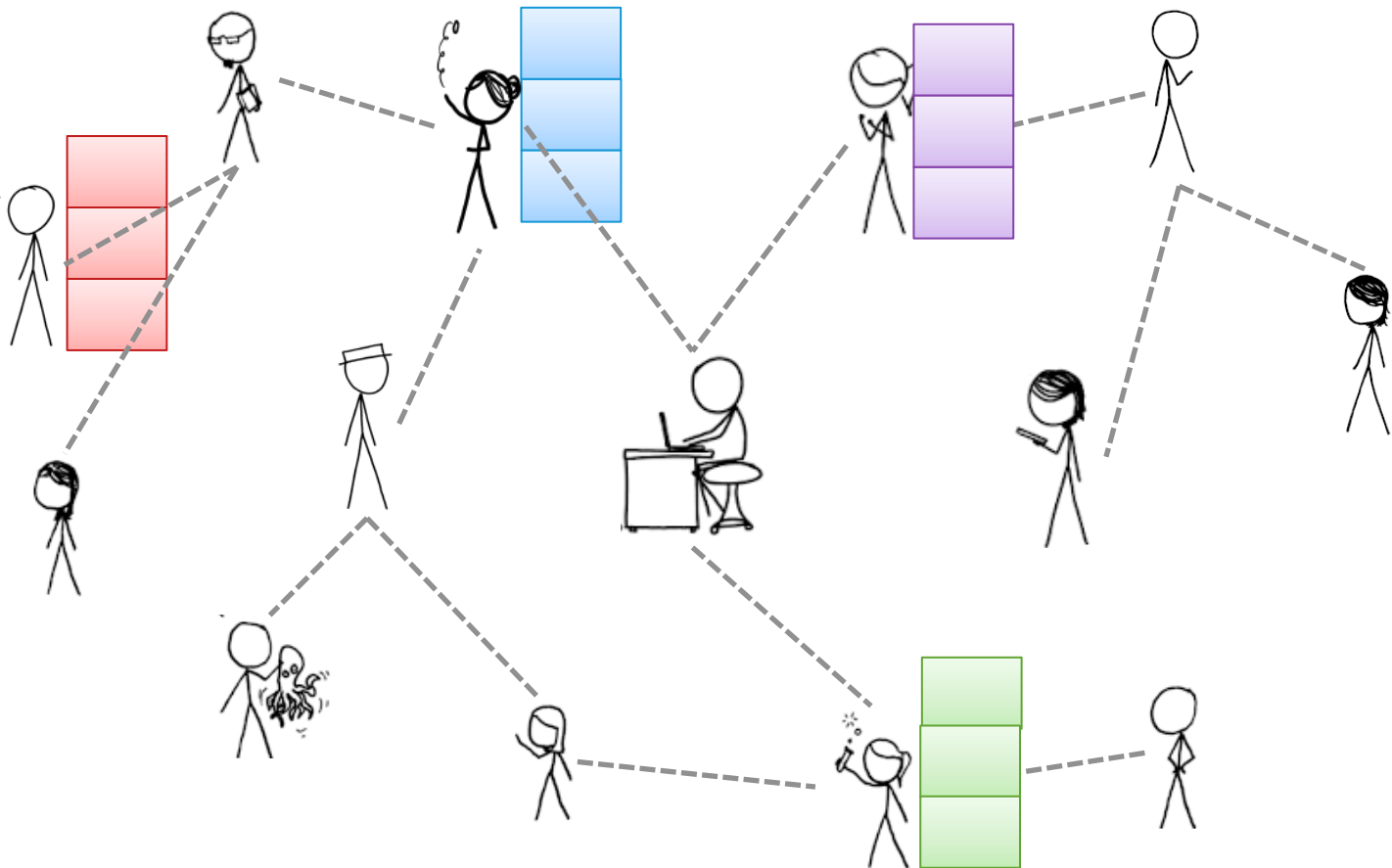
Types of Equilibria

- Equitable distribution of work



Types of Equilibria

- Specialized distribution of work



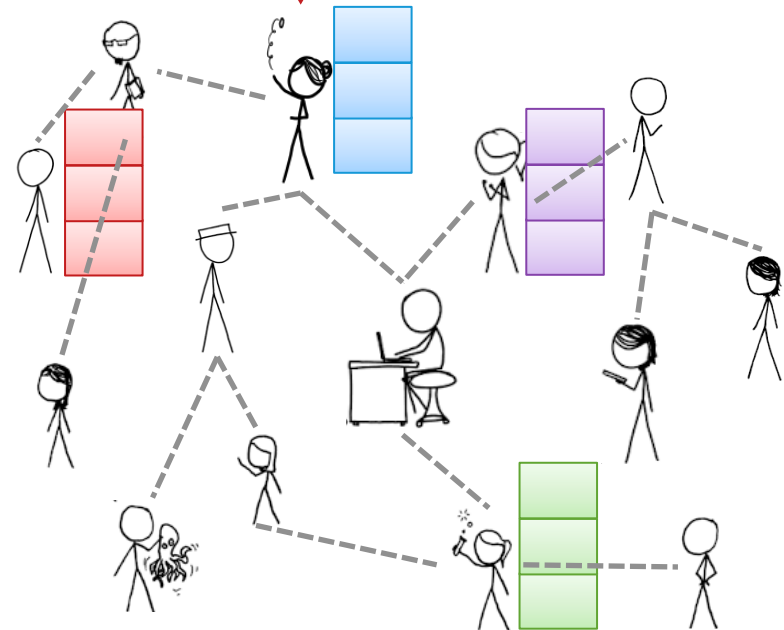
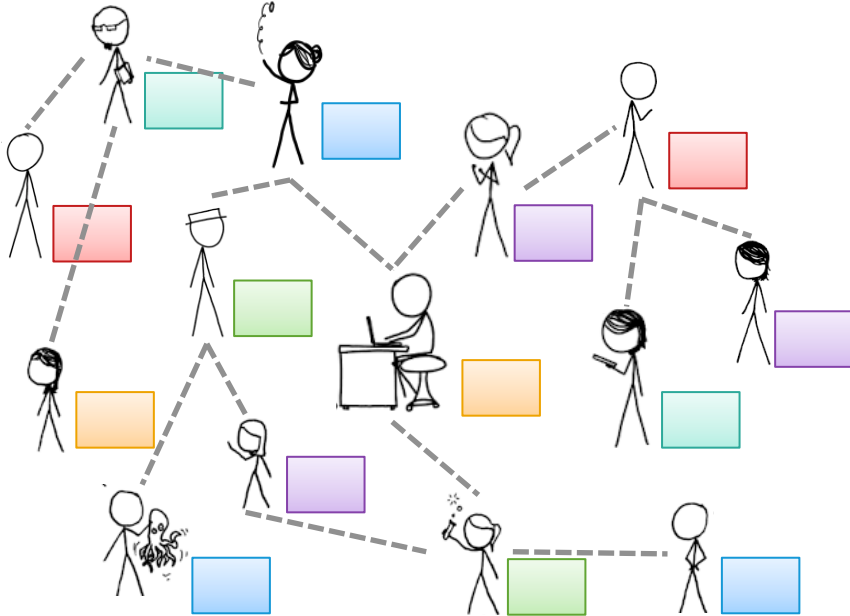
THM: Conditions for Specialization

$$\tau < \hat{\tau} = f(\lambda_{min})$$

yes

no

?

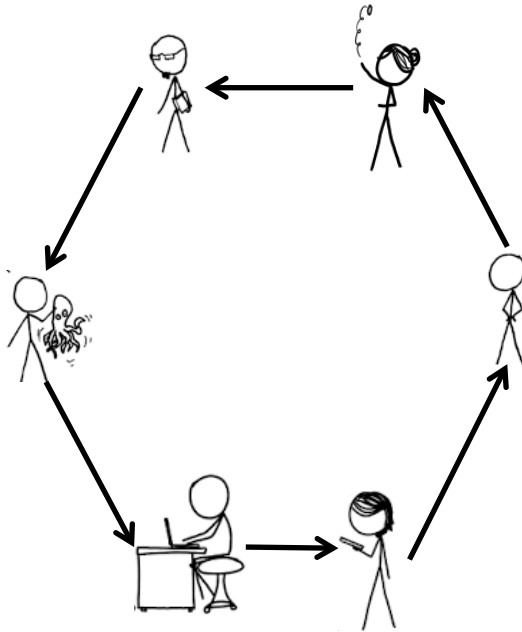


Smaller $\tau \implies$ less specialized equilibrium
i.e. shorter lived content is less specialized

Specialization occurs with
longer lived content

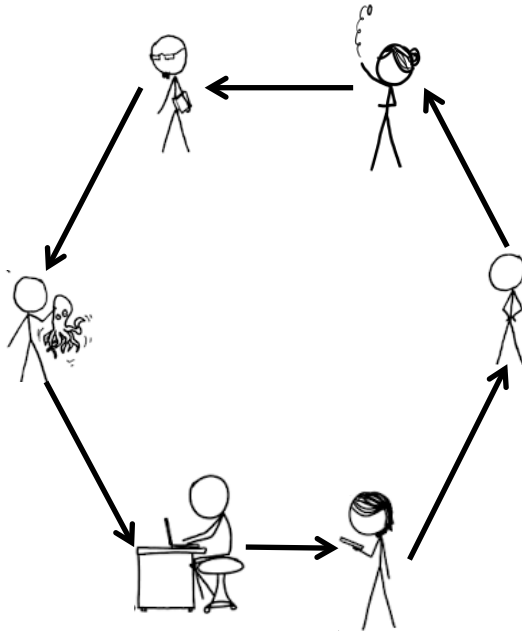
Specialization: Toy Example

- How much effort does each person expend?



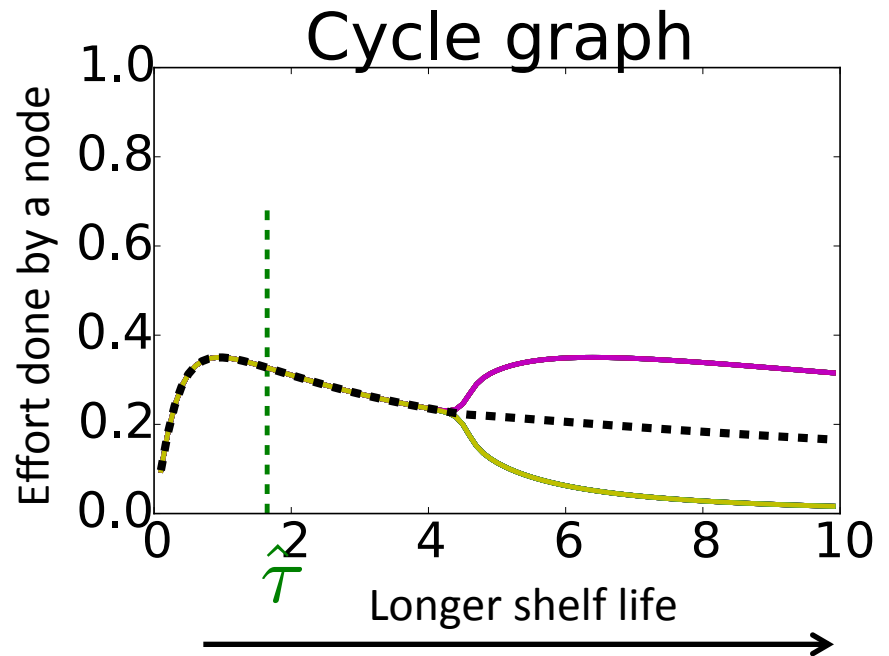
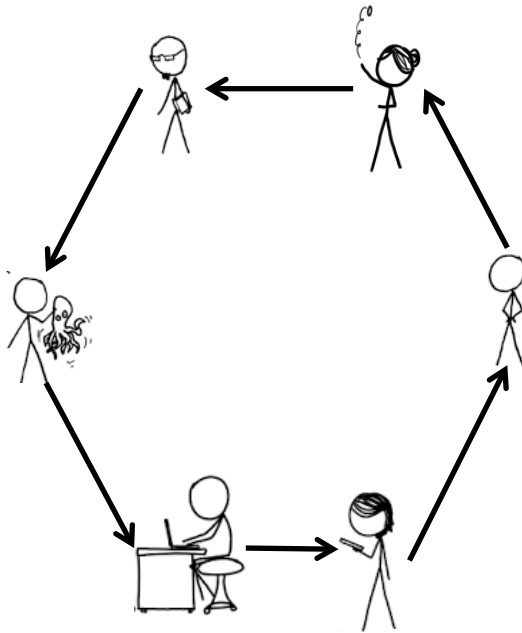
Specialization: Toy Example

Depends on \mathcal{T}

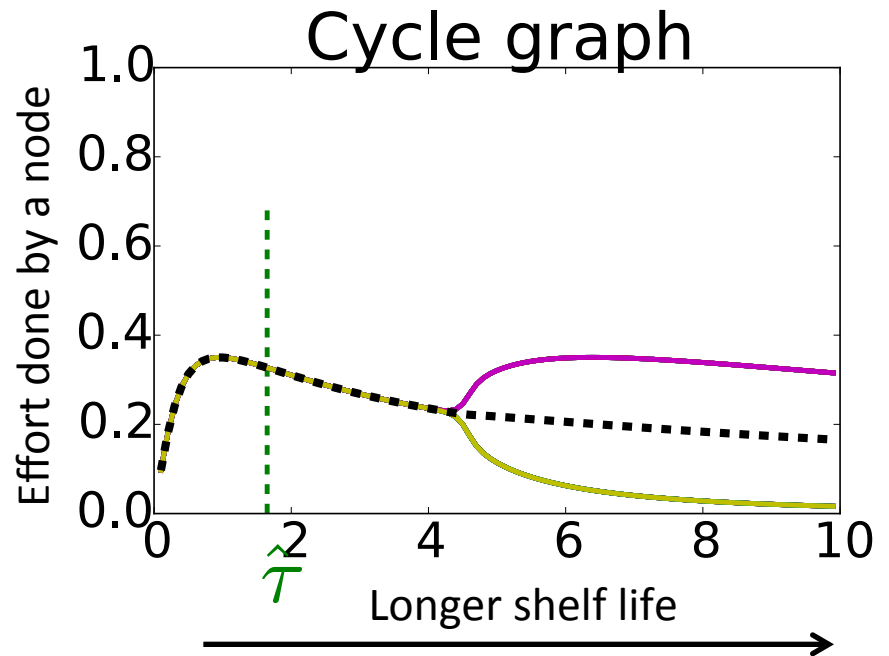
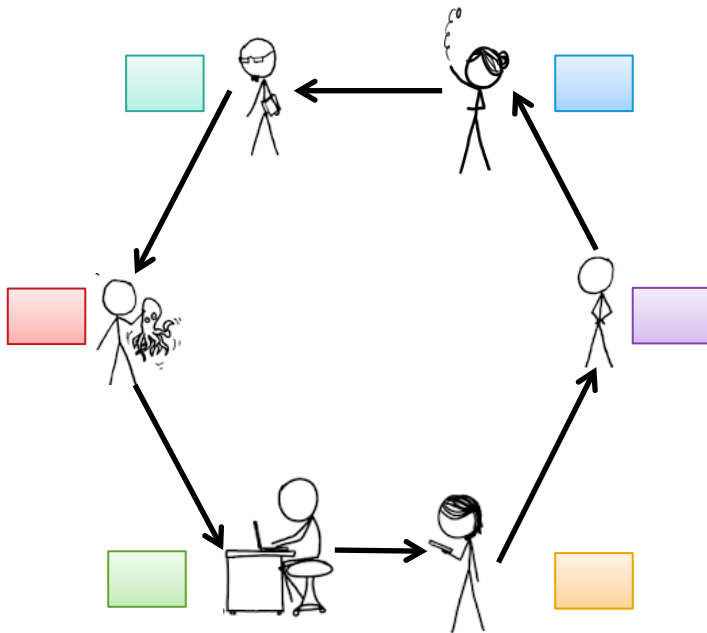


Specialization: Toy Example

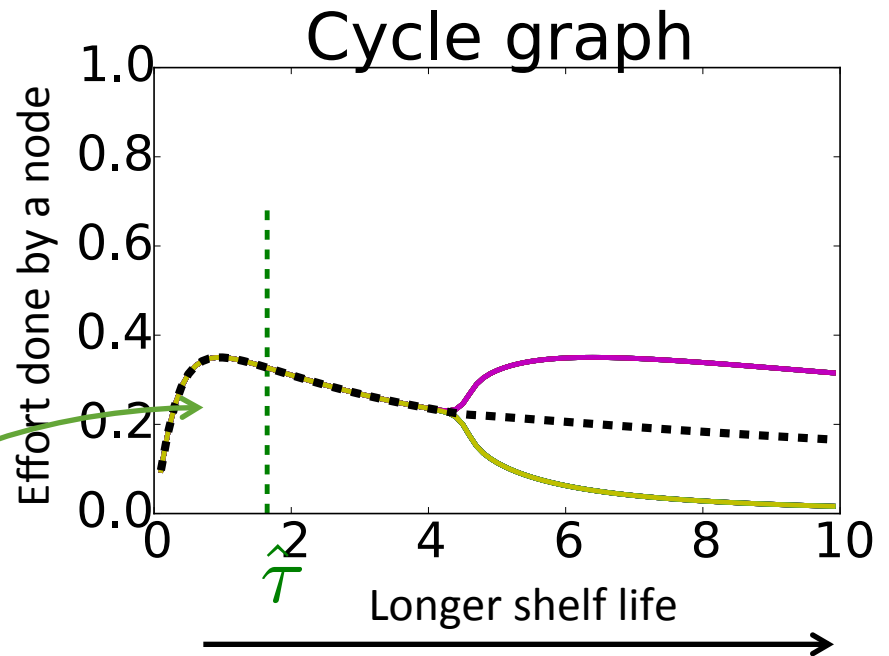
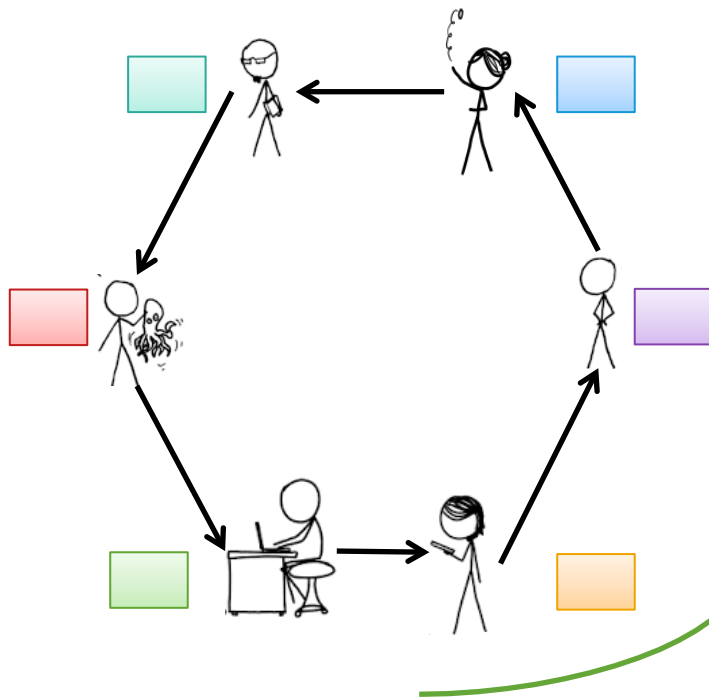
Depends on τ



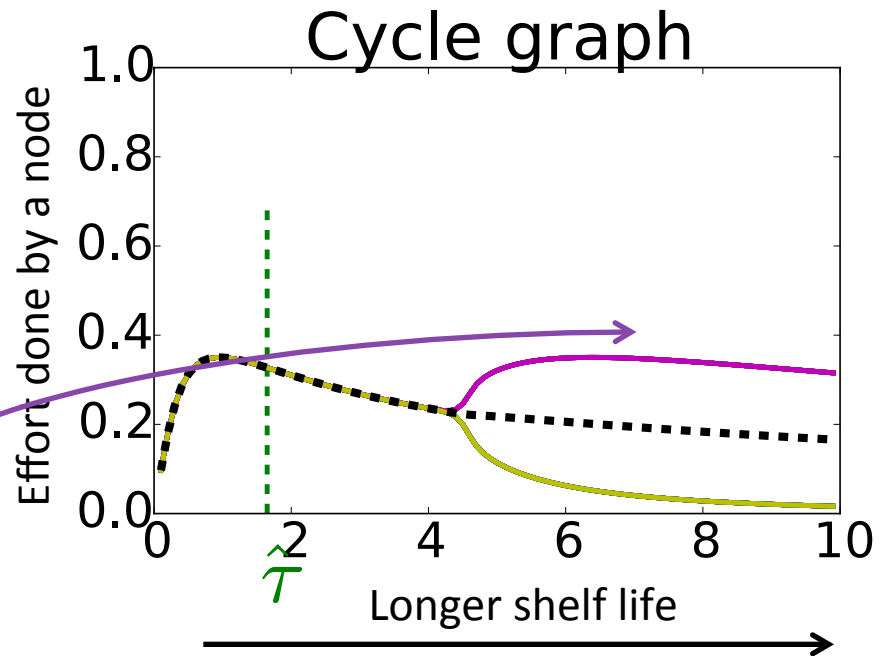
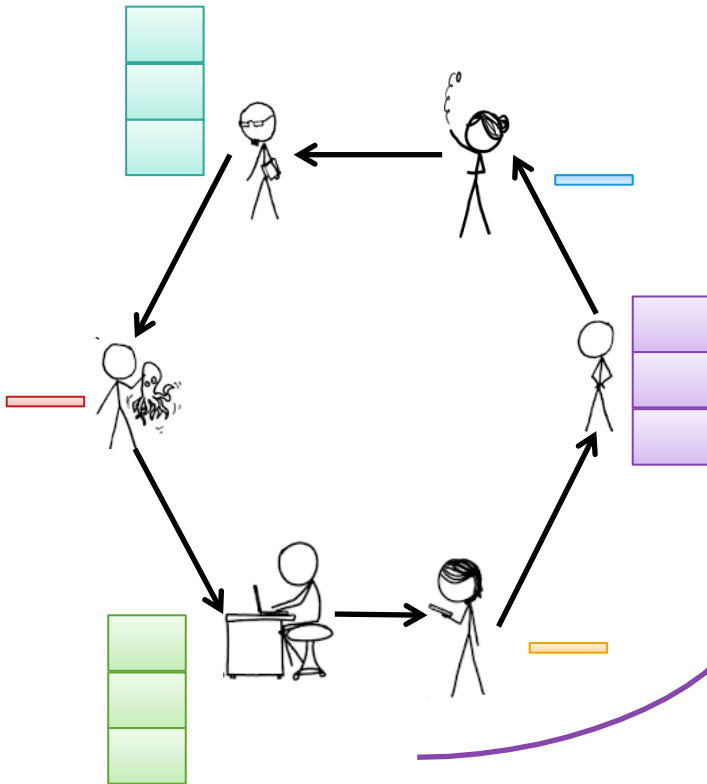
Specialization: Toy Example



Specialization: Toy Example



Specialization: Toy Example



Specialization in Different Graphs

Graph	λ_{min}	$\hat{\tau}$
Complete	-1	$\forall \tau$
Cycle (Even)	-2	\sqrt{e}
Cycle (Odd)	$-2 + \frac{\pi^2}{n^2}$	$\frac{n}{(n^2 - \pi^2)^{\frac{1}{2}}} e^{\frac{n^2}{2(n^2 - \pi^2)}}$
Erdős-Renyi	$-2\sqrt{np}$	$(\frac{1}{2\sqrt{np} - 1})^{\frac{1}{2}} e^{\frac{1}{2(2\sqrt{np} - 1)}}$
Star	$-\sqrt{n-1}$	$(\frac{1}{\sqrt{n-1} - 1})^{\frac{1}{2}} e^{\frac{1}{2(\sqrt{n-1} - 1)}}$
Complete Bipartite	$-\frac{n}{2}$	$(\frac{2}{n-2})^{\frac{1}{2}} e^{\frac{1}{n-2}}$

Conclusion

- Specialization occurs in social graphs
 - Long lived content exhibits specialization
 - Specialization *cannot* occur for arbitrarily short lived content
- The existence of specialized equilibria are based on properties of the graph

Thank you

Any Questions?



Microsoft®
Research

Best Response

$$U(y_i, y_{-i}) = 1 - e^{-\tau(y_i + y_{-i})} - c(y_i)$$

$$\text{where } (c : y_i \rightarrow \frac{\theta}{\alpha + 1} y_i^{\alpha+1}), \alpha > 0$$

$$\phi(y_{-i}, \tau) = \frac{\alpha}{\tau} W\left(\frac{\tau^{\frac{\alpha+1}{\alpha}}}{\alpha} e^{-\tau y_{-i}}\right)$$

Condition for Nash Equilibrium

$$\tau < \hat{\tau} \stackrel{\text{def}}{=} \left(\frac{\alpha}{-\lambda_{\min} - 1} \right)^{\frac{\alpha}{\alpha+1}} e^{\frac{\alpha}{(\alpha+1)(-\lambda_{\min} - 1)}}$$