The YouTube™ Social Network

Mirjam Wattenhofer[1], Roger Wattenhofer[2], **Zack Zhu[2]**
Google Zurich[1], ETH Zurich[2]

Sixth International AAAI Conference on Weblogs and Social Media
Dublin, June 7th, 2012

Google™

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

In 2011, >1 trillion views (140/person on Earth)

>60 hours of video is uploaded every minute

…thousands of full-length movies

>4 billion videos are viewed everyday

Every week, 100 million people take social actions on YouTube

Millions of subscriptions happen each day…

>800 million unique visitors on YouTube every month

>50% of videos have been rated/commented on

Source: www.youtube.com/t/press_statistics

In 2011, >1 trillion views (140/person on Earth)

>60 hours of video is uploaded every minute

…thousands of full-length movies

>4 billion videos are viewed everyday

Every week, 100 million people take social actions on YouTube

Millions of subscriptions happen each day…

>800 million unique visitors on YouTube every month

>50% of videos have been rated/commented on

Source: www.youtube.com/t/press_statistics

In 2011, >1 trillion views (140/person on Earth)

>60 hours of video is uploaded every minute

…thousands of full-length movies

>4 billion videos are viewed everyday

Every week, 100 million people take social actions on YouTube

Millions of subscriptions happen each day…

>800 million unique visitors on YouTube every month

>50% of videos have been rated/commented on

Source: www.youtube.com/t/press_statistics

In 2011, >1 trillion views (140/person on Earth)

>60 hours of video is uploaded every minute

…thousands of full-length movies

>4 billion videos are viewed everyday

Every week, 100 million people take social actions on YouTube

Millions of subscriptions happen each day…

>800 million unique visitors on YouTube every month

>50% of videos have been rated/commented on

Source: www.youtube.com/t/press_statistics

Questions…

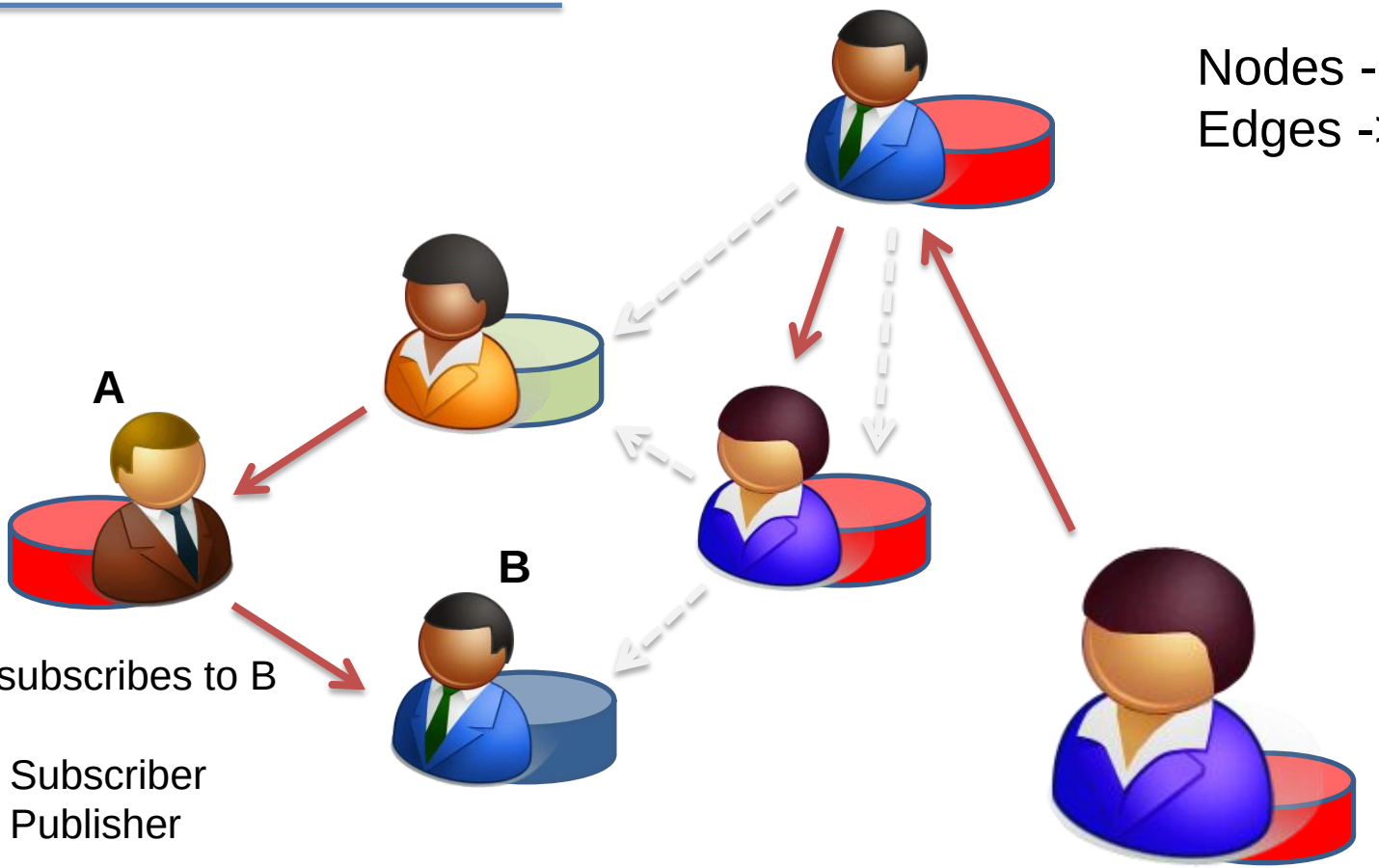
1. Are YouTube users behaving similarly to users of traditional social networks?
   - How do users interact and connect?

2. How are members of the YouTube Partner Program selected?
   - What is the interplay between content popularity and social popularity?
   - Which measures of popularity are indicative of potential partners?

YouTube &
Traditional Social Networks

Popularity &
YouTube Partners

Measurements

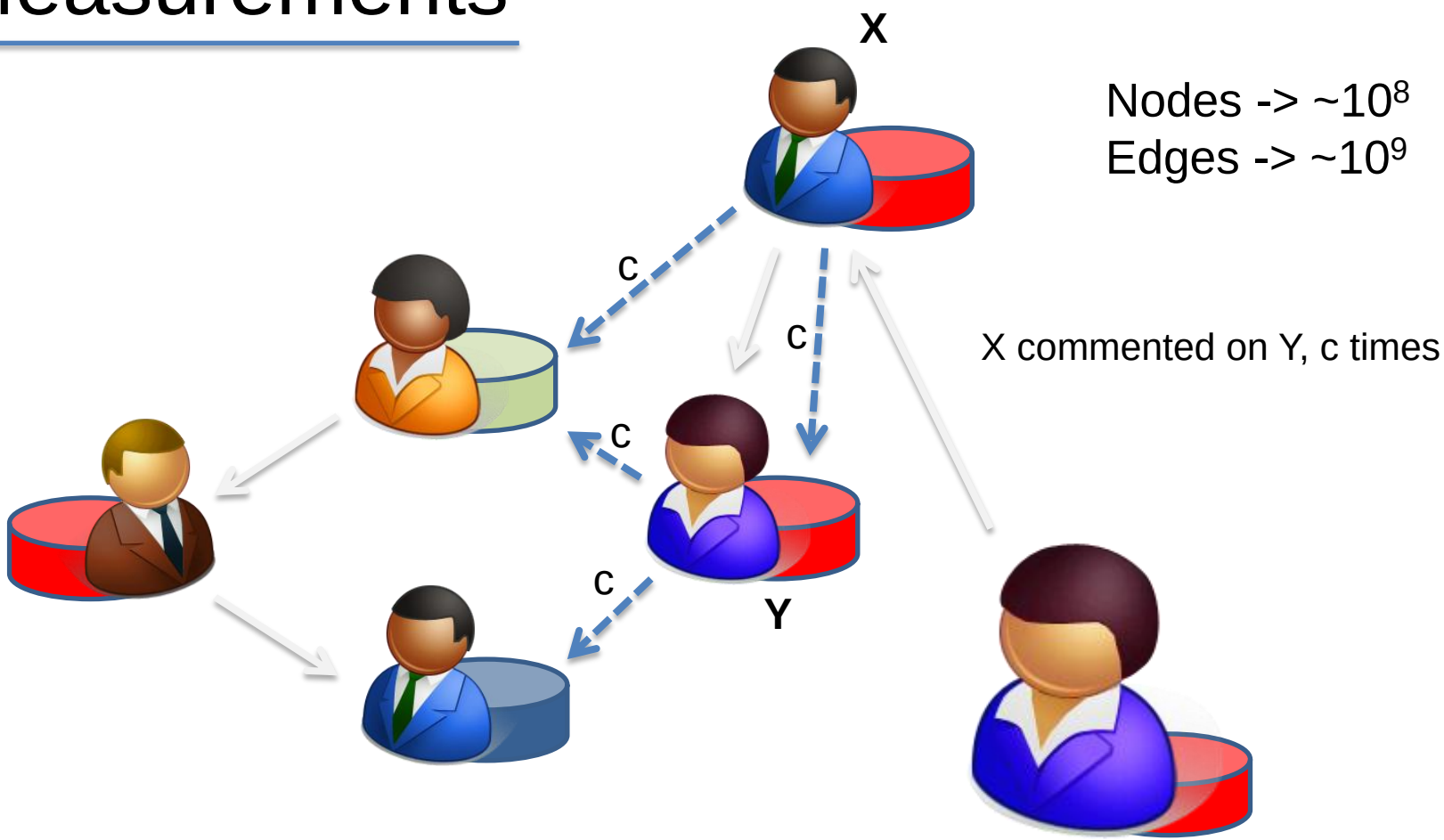


Nodes -> ~10^8
Edges -> ~10^9

A subscribes to B

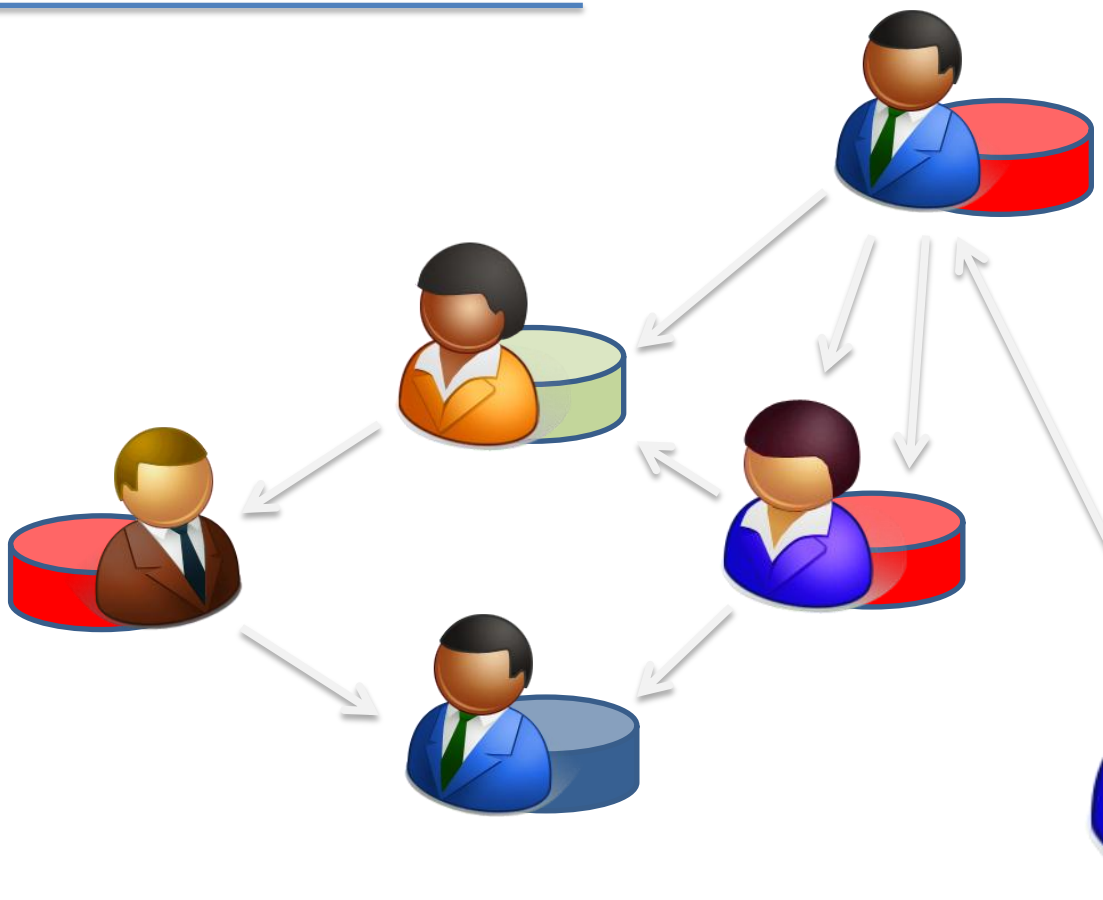
A: Subscriber
B: Publisher

Measurements



Nodes -> ~10^8
Edges -> ~10^9

Measurements



Social Features
Deg. on sub-graph
Deg. on com-graph
PageRanks
Reciprocal Links
.
.
.

Content features

.

Max/med/min video views
Mode upload category
Max/med/min avg. rating
Max/med/min rating quantity
Upload quantity

.
.
.

$\mathfrak{R}^{1\times 28}$

We measure the complete **subscription graph**, **comment graph** and **video metrics** aggregated to the user level. This snapshot was compiled in August 2011.

Connecting & Interacting

Given user μ, we compare sets

$$C_\mu \in \{c_1, c_2, \dots, c_n\}$$
$$S_\mu \in \{s_1, s_2, \dots, s_n\}$$

for the overlap proportion

$$\rho_\mu = \frac{S_\mu \cap C_\mu}{S_\mu \cup C_\mu}$$

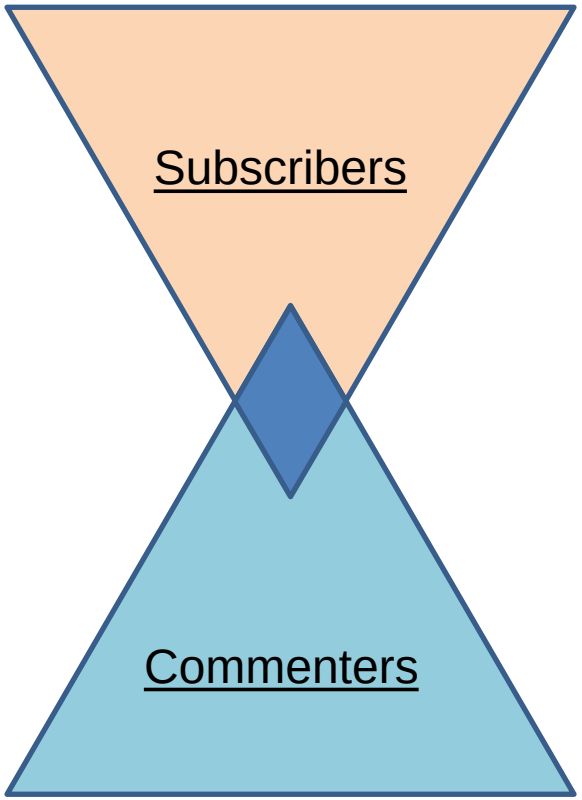
Connecting & Interacting

Given user μ, we compare sets

$$C_\mu \in \{c_1, c_2, \dots, c_n\}$$
$$S_\mu \in \{s_1, s_2, \dots, s_n\}$$

for the overlap proportion

$$\rho_\mu = \frac{S_\mu \cap C_\mu}{S_\mu \cup C_\mu}$$



Connecting & Interacting

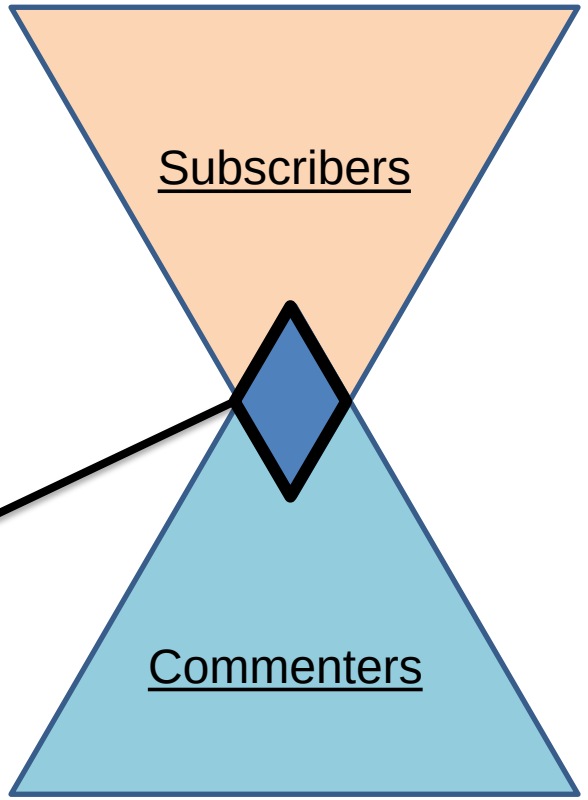
Given user μ, we compare sets

$$C_\mu \in \{c_1, c_2, \ldots, c_n\}$$
$$S_\mu \in \{s_1, s_2, \ldots, s_n\}$$

for the overlap proportion

$$\rho_\mu = \frac{S_\mu \cap C_\mu}{S_\mu \cup C_\mu}$$

On average: $\bar{\rho}$ **= 9.6%** *(s.d. 1.9%)*



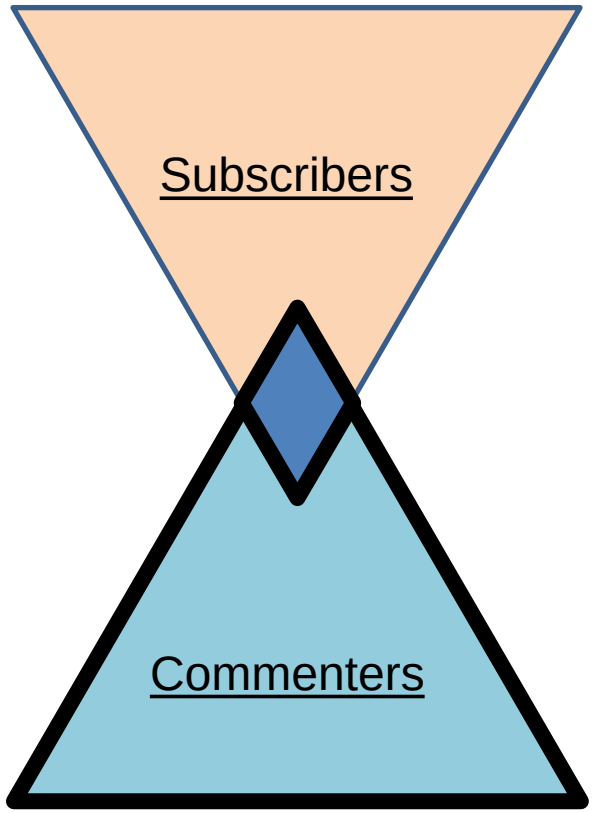
Connecting & Interacting

Given user μ, we compare sets

$$C_\mu \in \{c_1, c_2, \ldots, c_n\}$$
$$S_\mu \in \{s_1, s_2, \ldots, s_n\}$$

for the overlap proportion

$$\rho_\mu = \frac{S_\mu \cap C_\mu}{S_\mu \cup C_\mu}$$



On average: $\bar{\rho}$ **= 9.6%** *(s.d. 1.9%)*

$$\frac{1}{n}\sum_{\mu \in N} \frac{S_\mu \cap C_\mu}{C_\mu} = 16.46\%$$

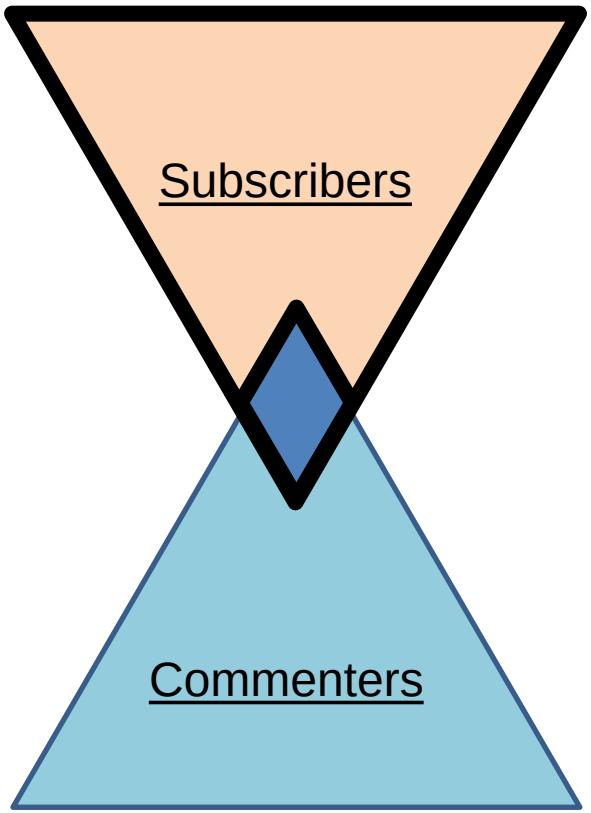
Connecting & Interacting

Given user μ, we compare sets

$$C_\mu \in \{c_1, c_2, \ldots, c_n\}$$
$$S_\mu \in \{s_1, s_2, \ldots, s_n\}$$

for the overlap proportion

$$\rho_\mu = \frac{S_\mu \cap C_\mu}{S_\mu \cup C_\mu}$$



On average: $\bar{\rho}$ **= 9.6%** *(s.d. 1.9%)*

$$\frac{1}{n}\sum_{\mu \in N} \frac{S_\mu \cap C_\mu}{C_\mu} = 16.46\% \qquad \frac{1}{n}\sum_{\mu \in N} \frac{S_\mu \cap C_\mu}{S_\mu} = 18.1\%$$

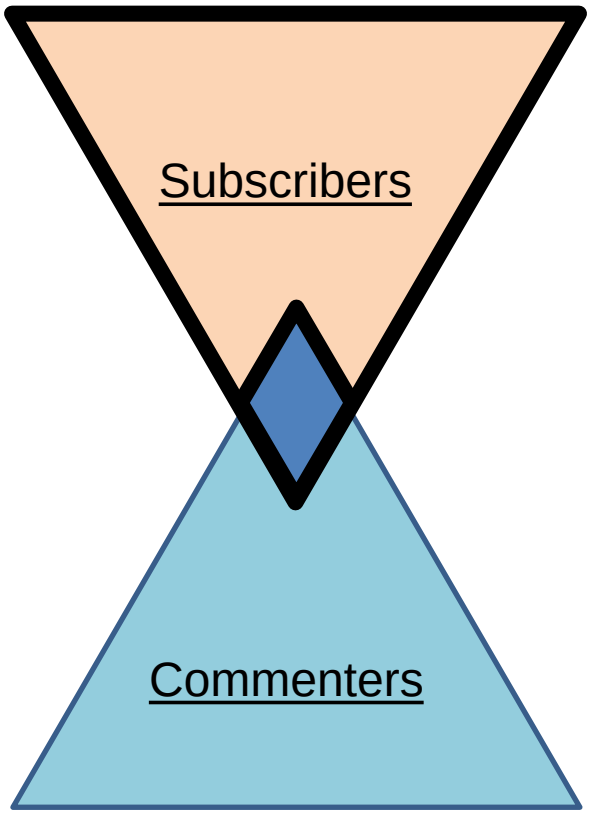
Connecting & Interacting

Given user μ, we compare sets

$$C_\mu \in \{c_1, c_2, \ldots, c_n\}$$
$$S_\mu \in \{s_1, s_2, \ldots, s_n\}$$

for the overlap proportion

$$\rho_\mu = \frac{S_\mu \cap C_\mu}{S_\mu \cup C_\mu}$$

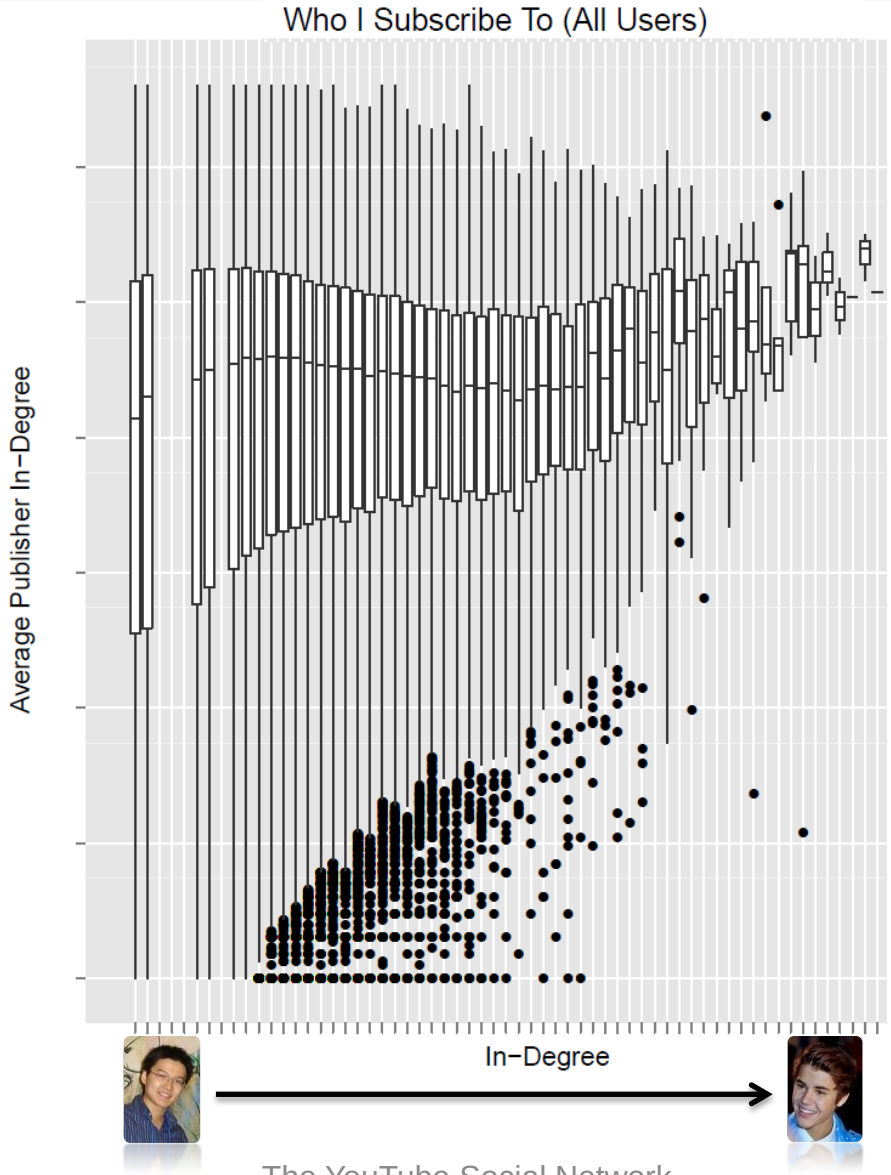


On average: $\bar{\rho}$ **= 9.6%** *(s.d. 1.9%)*

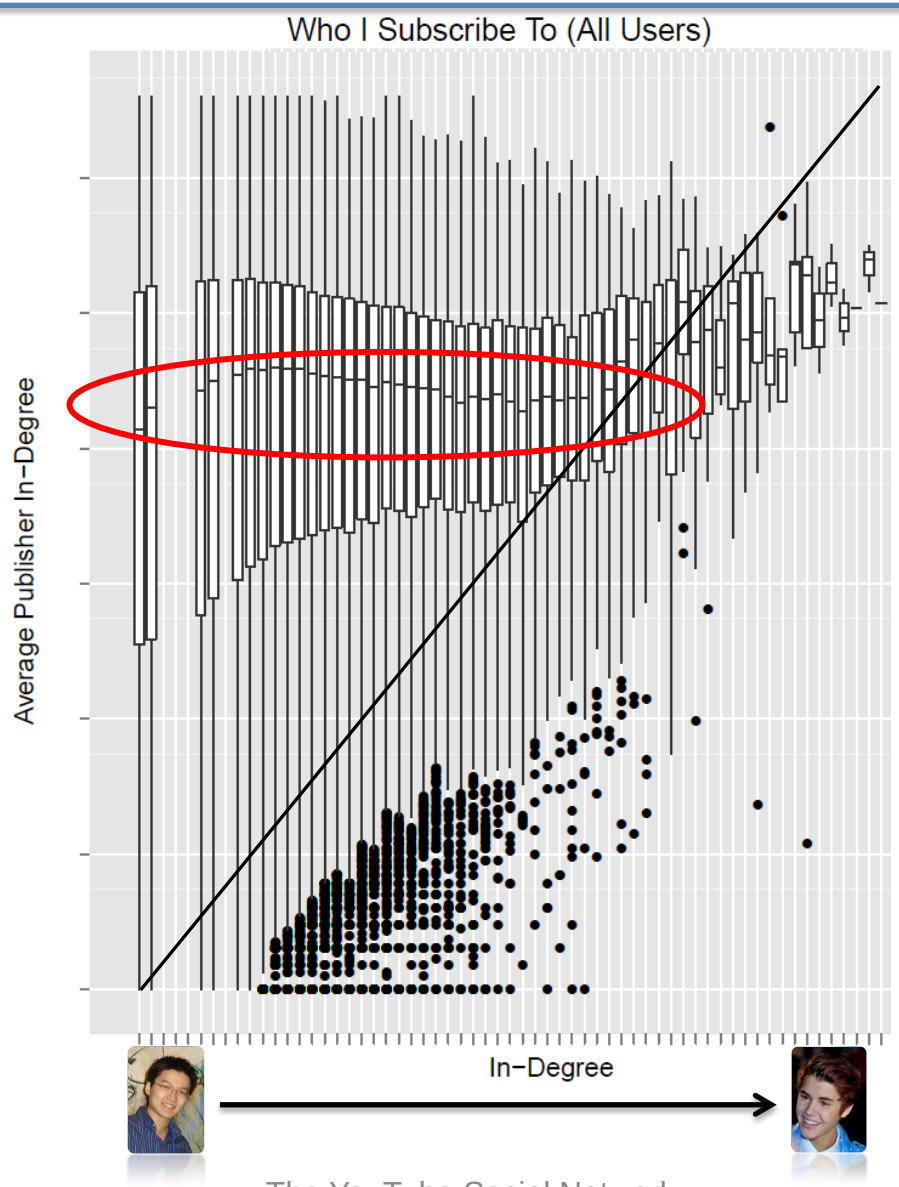
$$\frac{1}{n}\sum_{\mu \in N} \frac{S_\mu \cap C_\mu}{C_u} = 16.46\% \qquad \frac{1}{n}\sum_{\mu \in N} \frac{S_\mu \cap C_\mu}{S_u} = 18.1\%$$

Only a small portion of a user's subscribers are also commenters and vice-versa. The set of commenters and subscribers are **highly disparate populations**

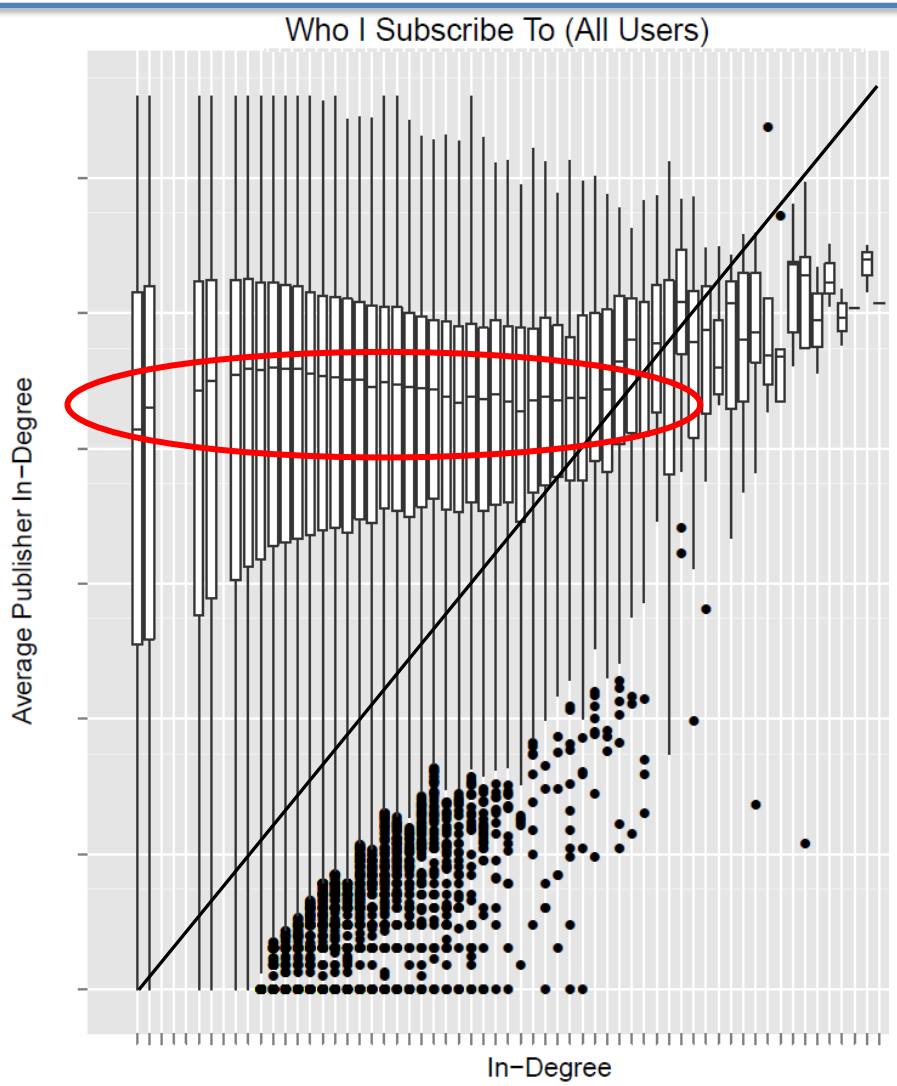
Assortative Linking



Assortative Linking



~~Assortative Linking~~

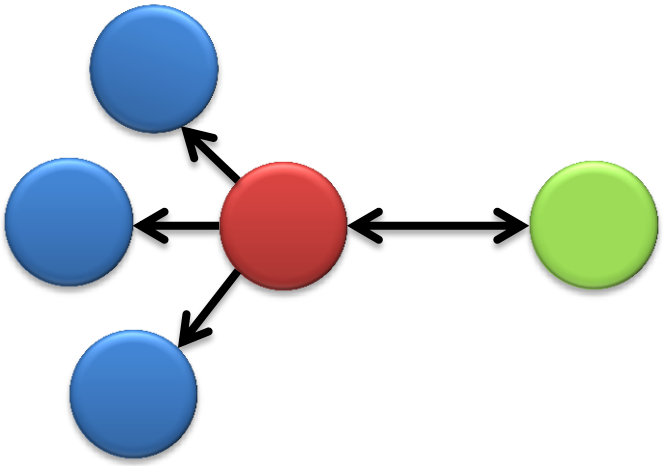


YouTube users **link disassortatively**, such that most subscribe to power-users.

Reciprocity

- Measurements of flickr and YAHOO! 360° BETA illustrate the presence of reciprocal linking amongst users *[Kumar et. al. 2006, Cha et. al. 2009]*

- We measure % of users with ≥ 1 reciprocal link:

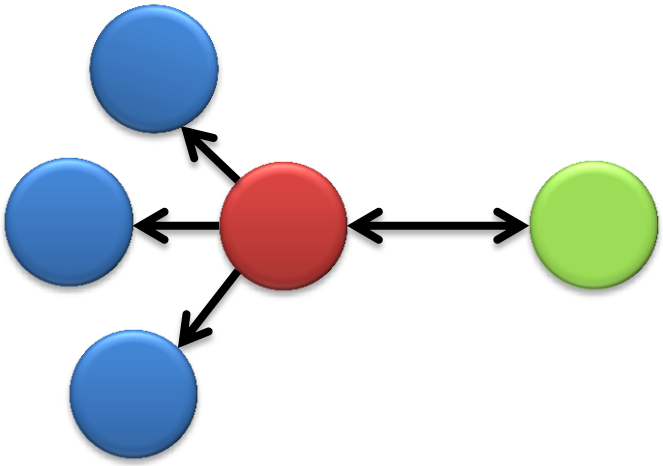


→ **~75%** have 0 reciprocal link

~~Reciprocity~~

- Measurements of flickr and YAHOO! 360° BETA illustrate the presence of reciprocal linking amongst users *[Kumar et. al. 2006, Cha et. al. 2009]*

- We measure % of users with ≥ 1 reciprocal link:



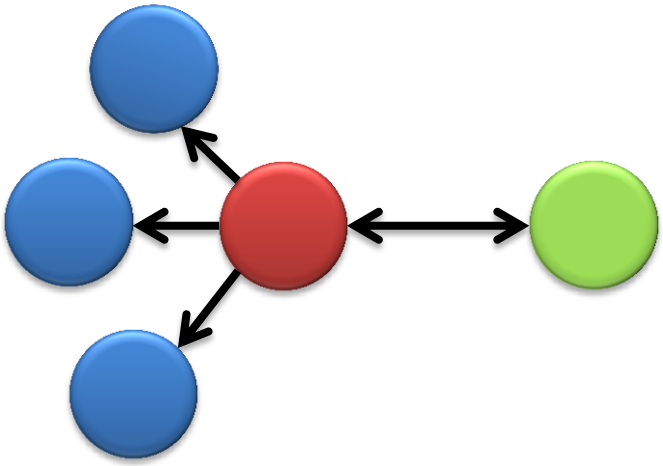
→ **~75%** have 0 reciprocal link

Approximately ¾ of the users have no reciprocal links, indicating **a lack of reciprocity** on the YouTube subscription graph.

~~Reciprocity~~

- Measurements of flickr and YAHOO! 360° BETA illustrate the presence of reciprocal linking amongst users *[Kumar et. al. 2006, Cha et. al. 2009]*

- We measure % of users with ≥ 1 reciprocal link:



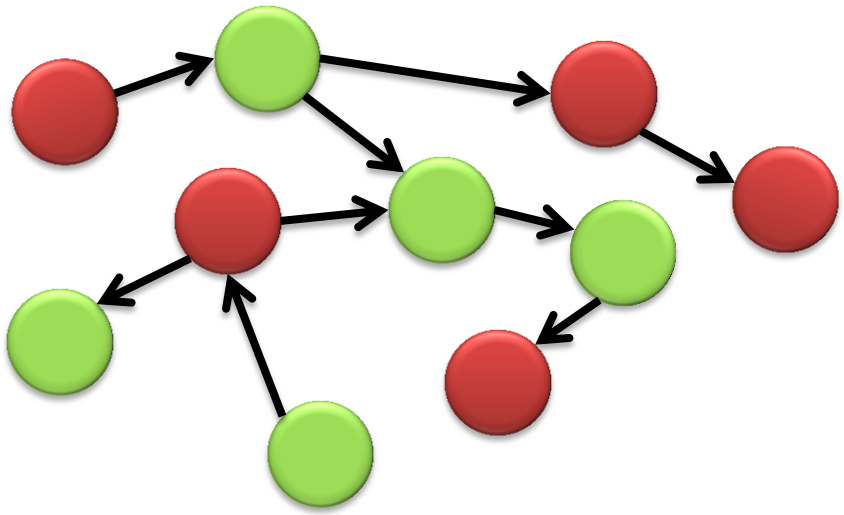
→ **~75%** have 0 reciprocal link

twitter

77.9% pairs are not linked reciprocally *[Kwak. et. al. 2010]*

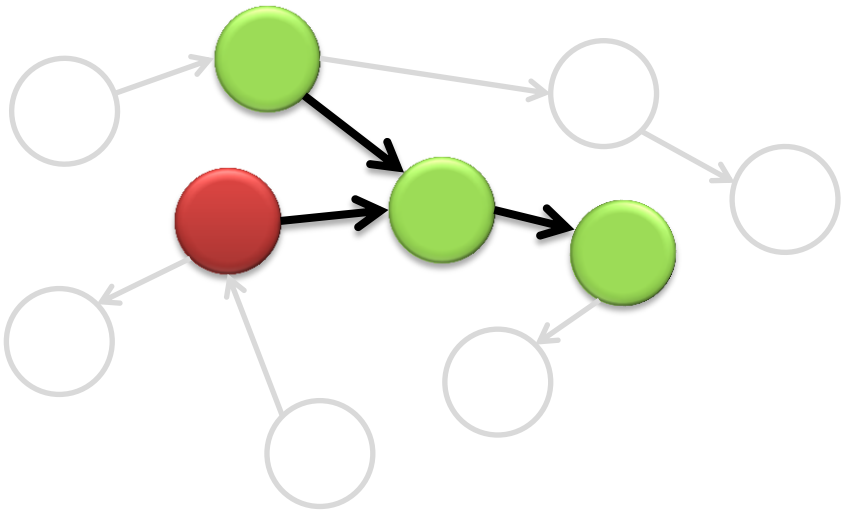
Approximately ¾ of the users have no reciprocal links, indicating **a lack of reciprocity** on the YouTube subscription graph.

Homophily



- Established by *[McPherson et. al., 2001]*
- Measure homophily via upload category
  - Hard assignment of all videos to 1 of 12 categories
  - Capture a **mode** upload category for each user
  - Use it as a proxy measure for user interest
- Compare whether a pair of linked users have the same mode category

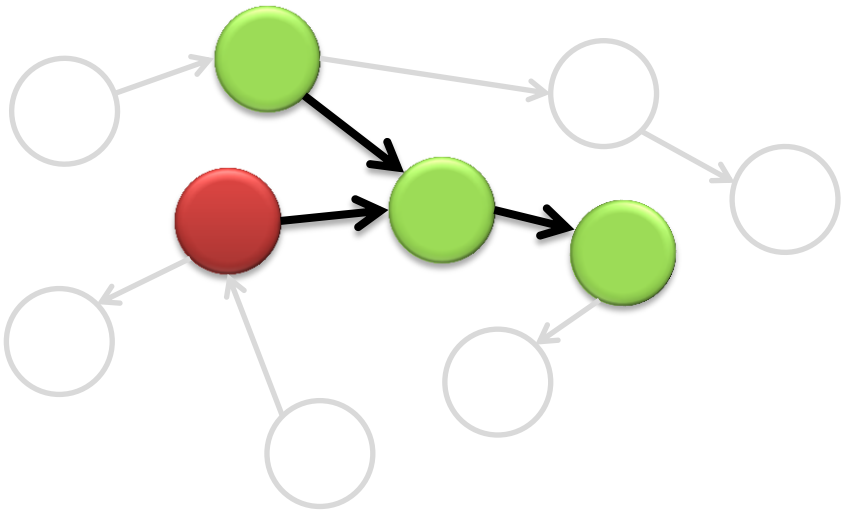
Homophily



Only **12.49%** of users are **more** connected with the same upload category on the subscription graph

- Established by *[McPherson et. al., 2001]*
- Measure homophily via upload category
  - Hard assignment of all videos to 1 of 12 categories
  - Capture a **mode** upload category for each user
  - Use it as a proxy measure for user interest
- Compare whether a pair of linked users have the same mode category

~~Homophily~~

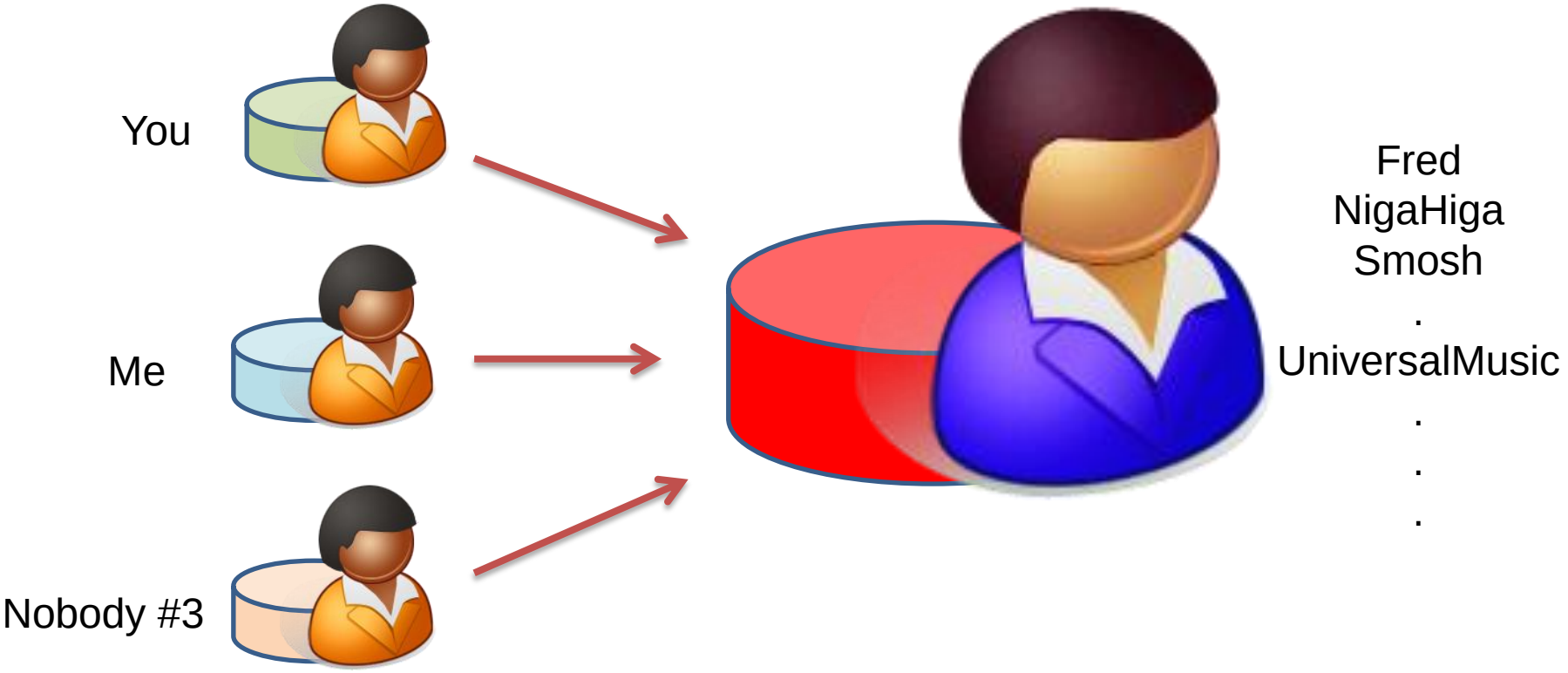


Only **12.49%** of users are **more** connected with the same upload category on the subscription graph

- Established by *[McPherson et. al., 2001]*
- Measure homophily via upload category
  - Hard assignment of all videos to 1 of 12 categories
  - Capture a **mode** upload category for each user
  - Use it as a proxy measure for user interest
- Compare whether a pair of linked users have the same mode category

By analyzing mode upload category, we observe a **lack of homophily** on both the subscription and comment graphs. Concurs with *[Kwak et. al., 2010]*'s study of Twitter.

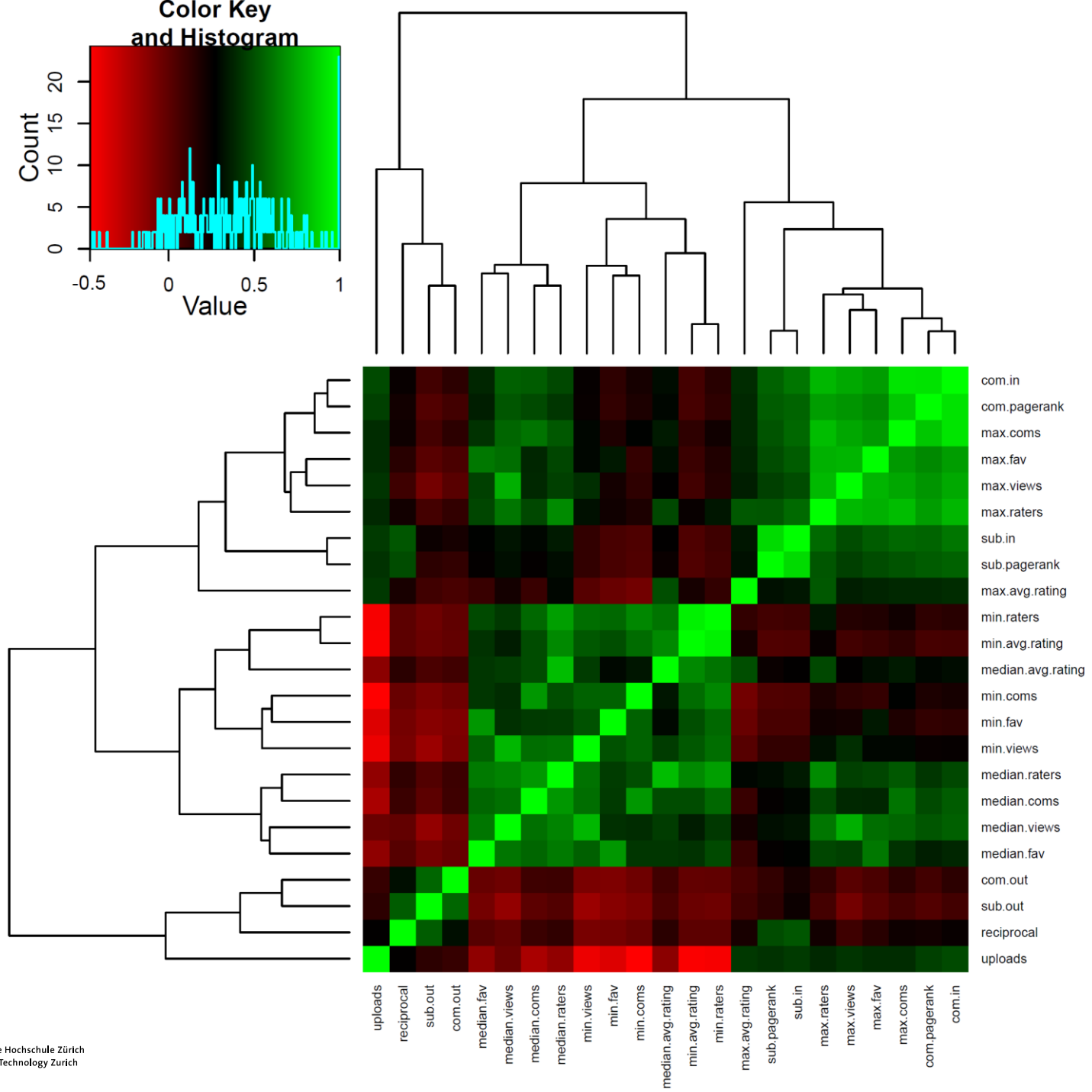
Social on YouTube



Social on YouTube stems from a relationship of **information generation and consumption** as opposed to real-world social relationships

YouTube &
Traditional Social Networks

Popularity &
YouTube Partners

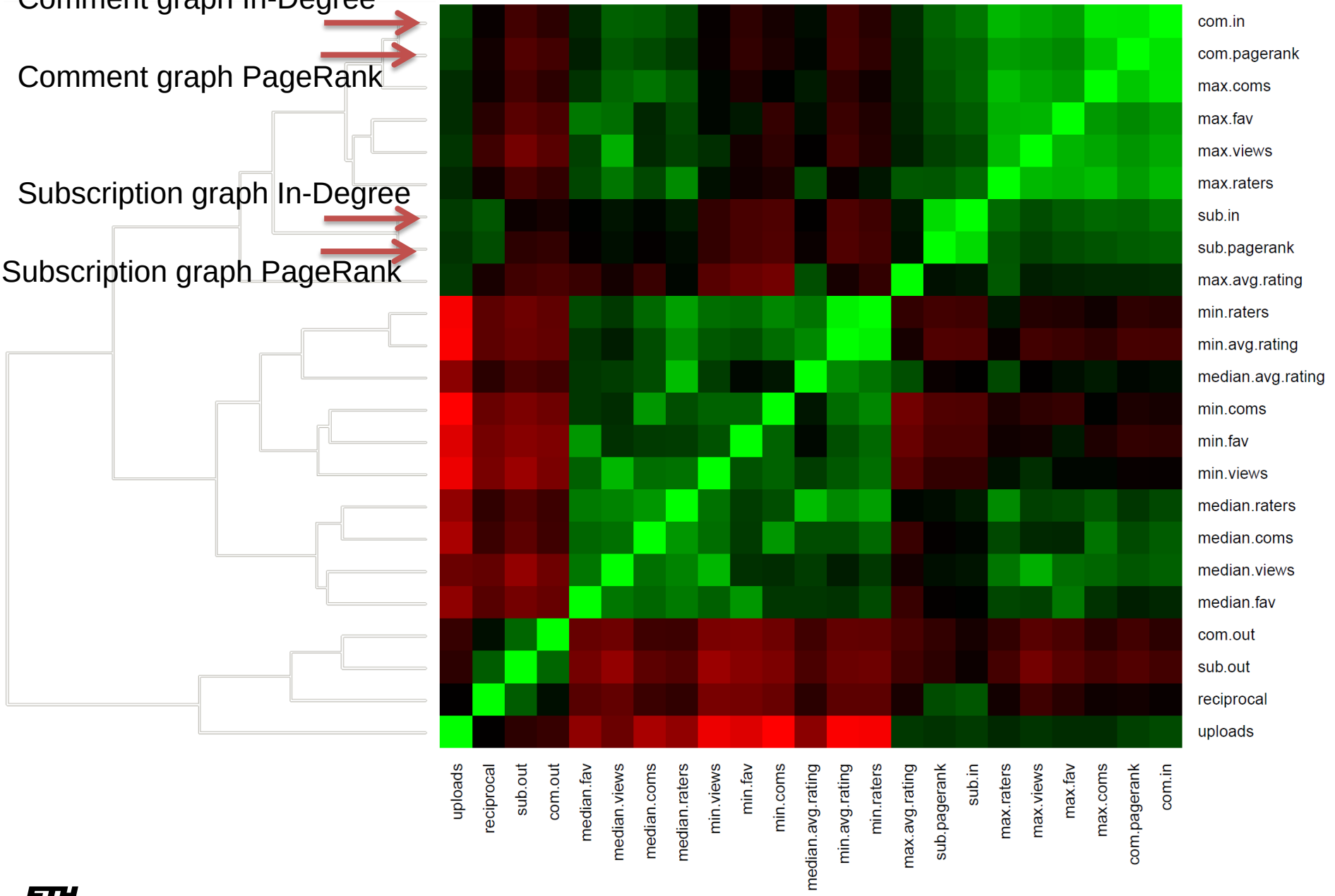

Color Key
and Histogram
Count
0
5
10
15
20
-0.5
0
0.5
1
Value
com.in
com.pagerank
max.coms
max.fav
max.views
max.raters
sub.in
sub.pagerank
max.avg.rating
min.raters
min.avg.rating
median.avg.rating
min.coms
min.fav
min.views
median.raters
median.coms
median.views
median.fav
com.out
sub.out
reciprocal
uploads
uploads
reciprocal
sub.out
com.out
median.fav
median.views
median.coms
median.raters
min.views
min.fav
min.coms
median.avg.rating
min.avg.rating
min.raters
max.avg.rating
sub.pagerank
sub.in
max.raters
max.views
max.fav
max.coms
com.pagerank
com.in

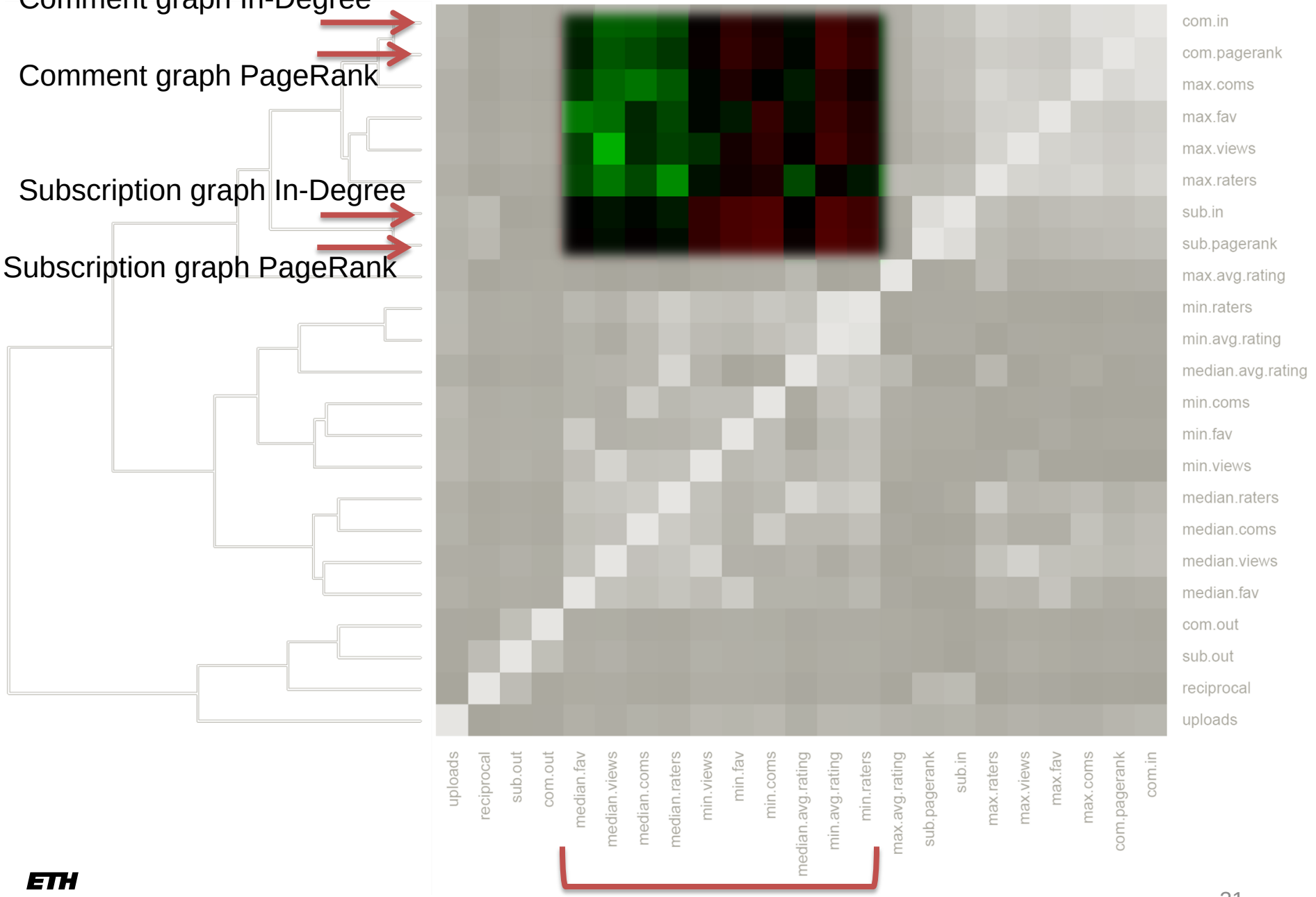
Comment graph In-Degree

Comment graph PageRank

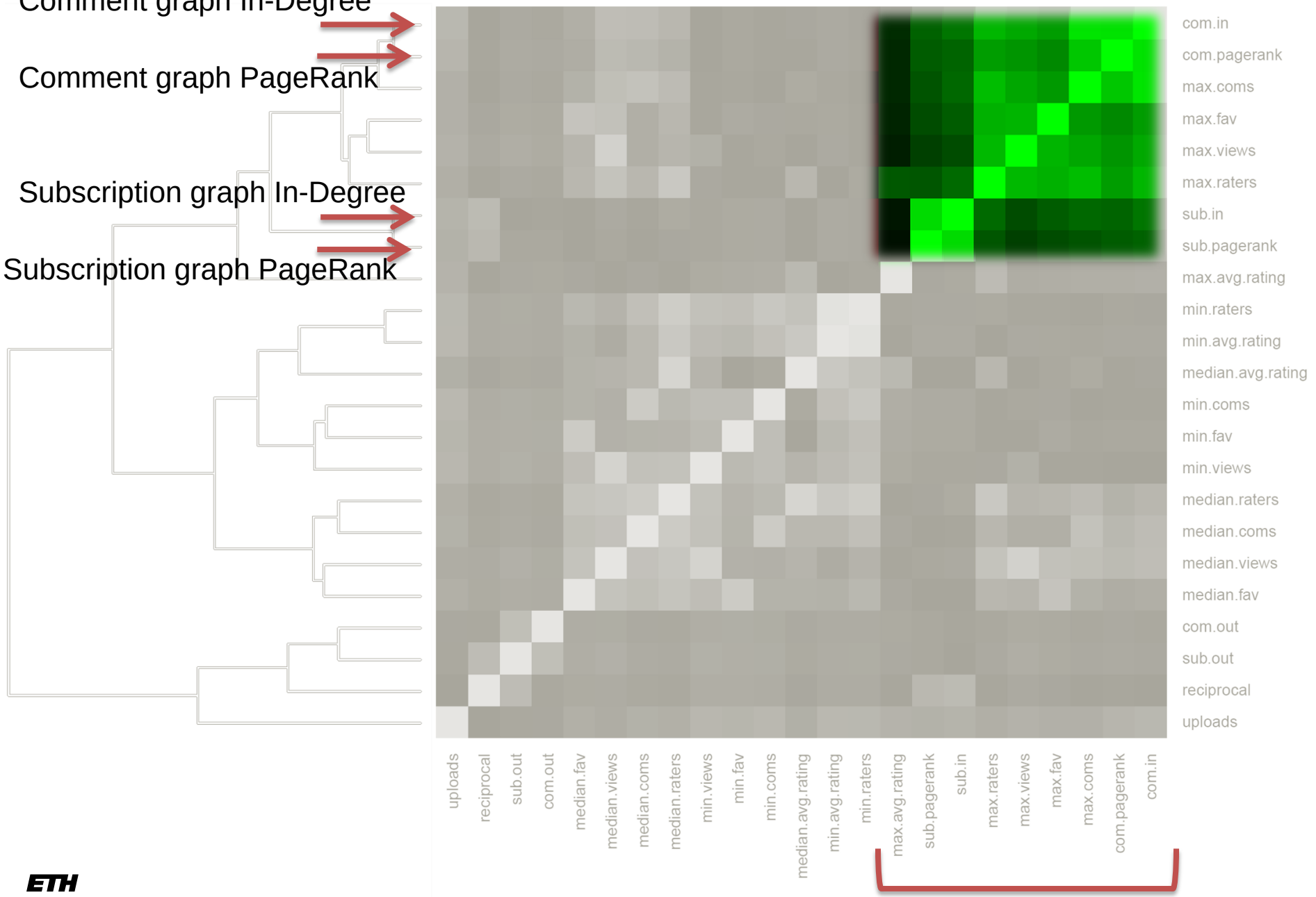
Subscription graph In-Degree

Subscription graph PageRank





Medians and minimums of content popularity


Comment graph In-Degree
Comment graph PageRank
Subscription graph In-Degree
Subscription graph PageRank
com.in
com.pagerank
max.coms
max.fav
max.views
max.raters
sub.in
sub.pagerank
max.avg.rating
min.raters
min.avg.rating
median.avg.rating
min.coms
min.fav
min.views
median.raters
median.coms
median.views
median.fav
com.out
sub.out
reciprocal
uploads
uploads
reciprocal
sub.out
com.out
median.fav
median.views
median.coms
median.raters
min.views
min.fav
min.coms
median.avg.rating
min.avg.rating
min.raters
max.avg.rating
sub.pagerank
sub.in
max.raters
max.views
max.fav
max.coms
com.pagerank
com.in
Maximums of content popularity

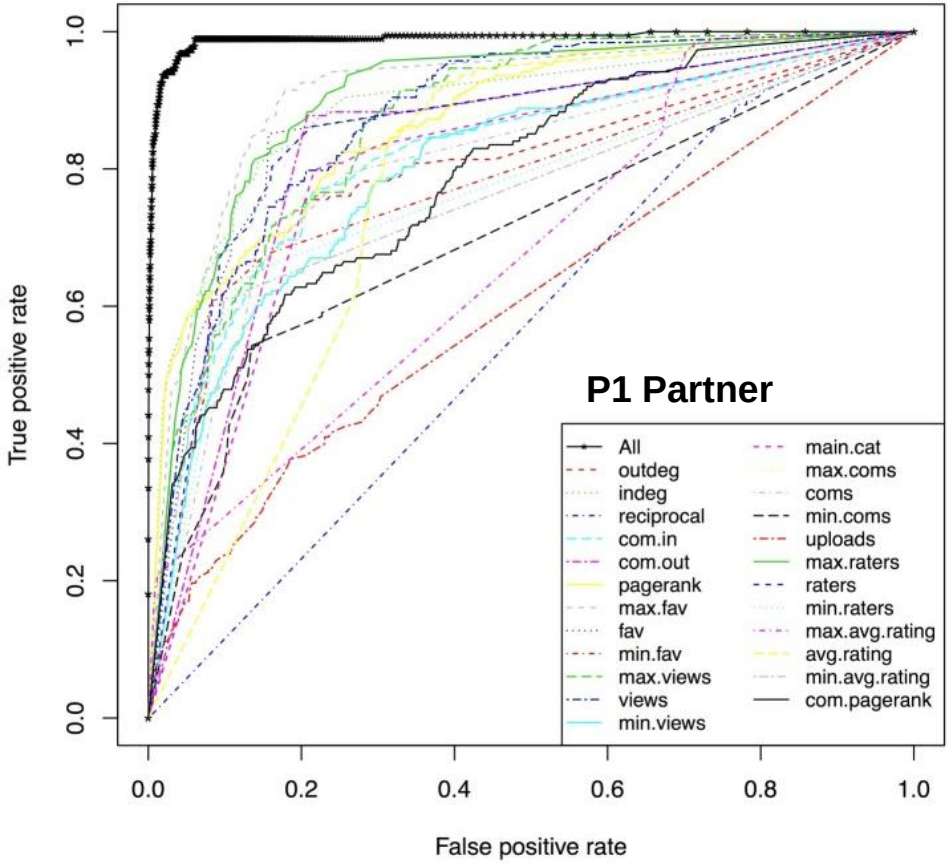
YouTube Partners

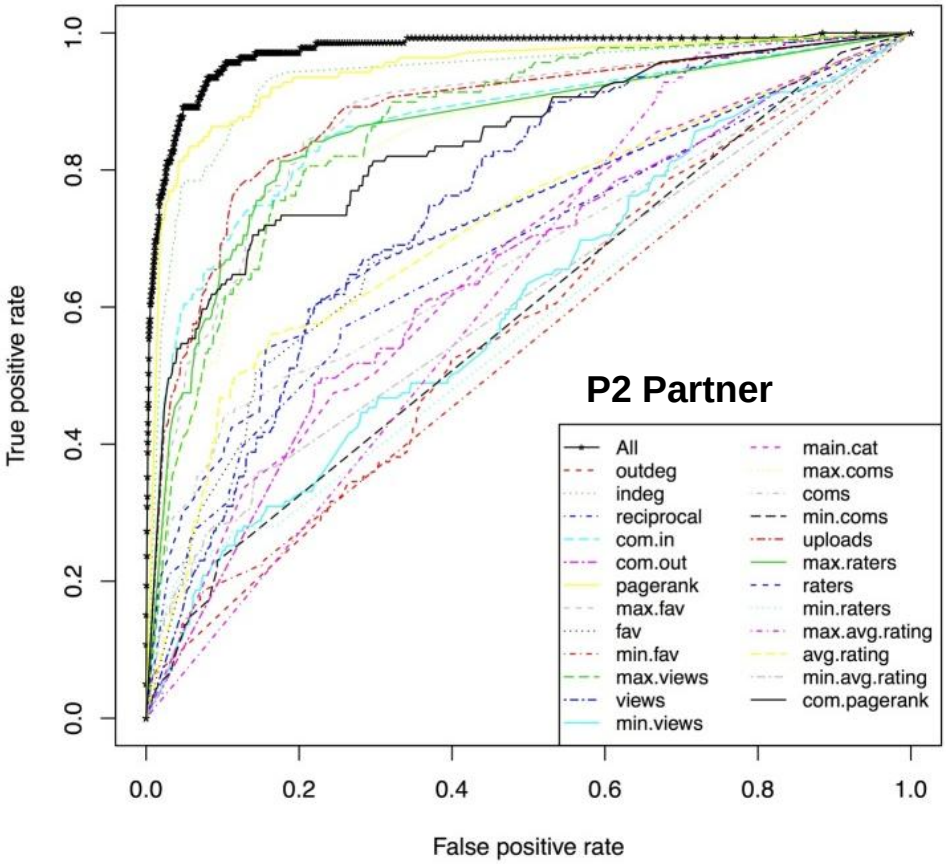
- YouTube Partner Program (YPP)[1]:
  - Created in 2007, top content contributors are selected to share ad revenue
  - Payout in the millions/year, top partners can make more than 100k/year
  - 30,000 partners from 27 countries

- Classify users as potential YPP members to aid the filtering process
  - Leverage **subscription graph**, **comment graph**, and **content metrics** as signals
  - Large-scale spotting problem
  - Focus on recall as opposed to precision

- Supervised-learning via Random Forest *[Liaw and Wiener, 2002]*
  - Three types of partners, formed as three **independent binary classification** problem
  - Feature selection via Gini entropy reduction

[1]www.youtube.com/t/press_statistics

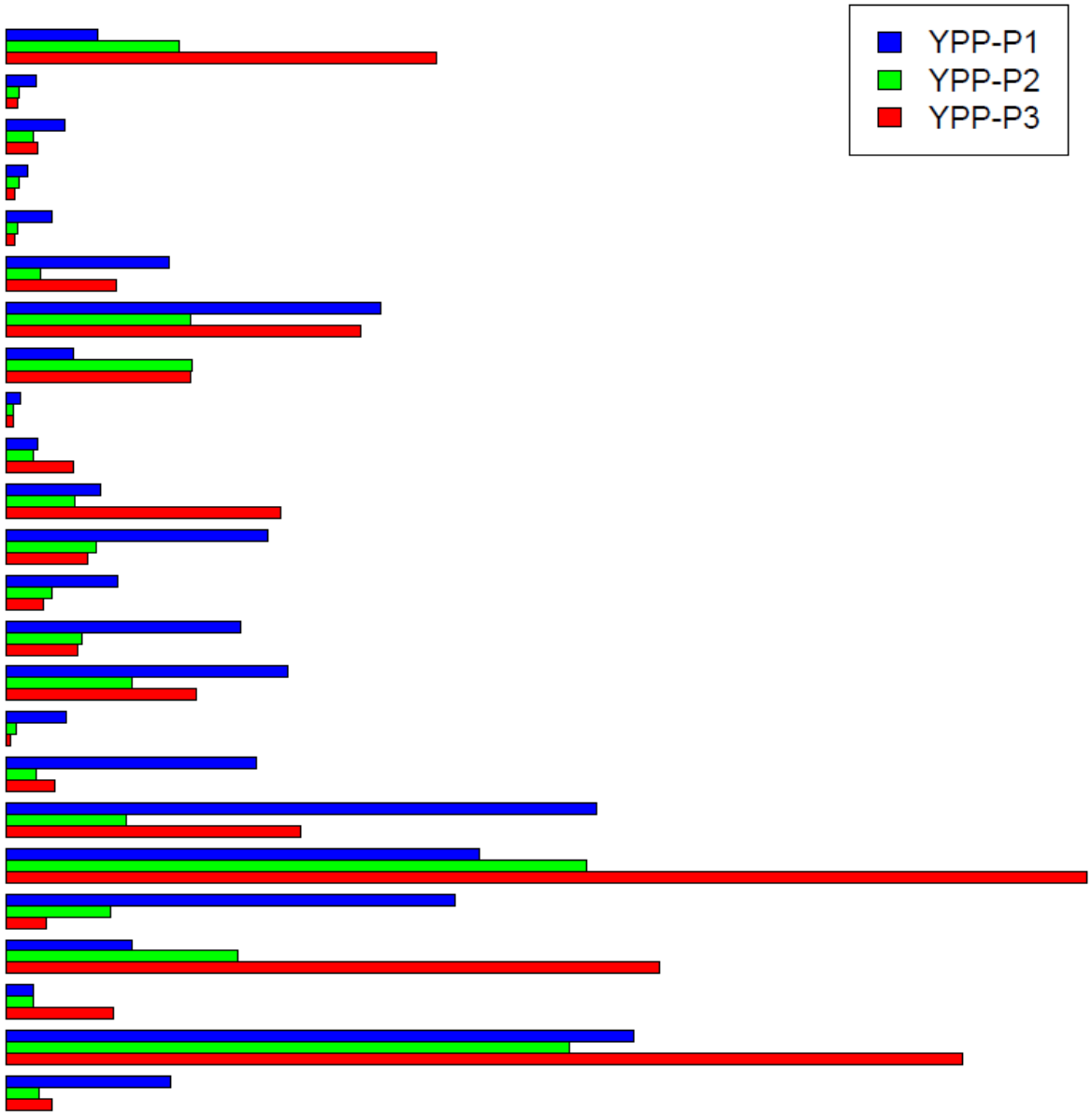




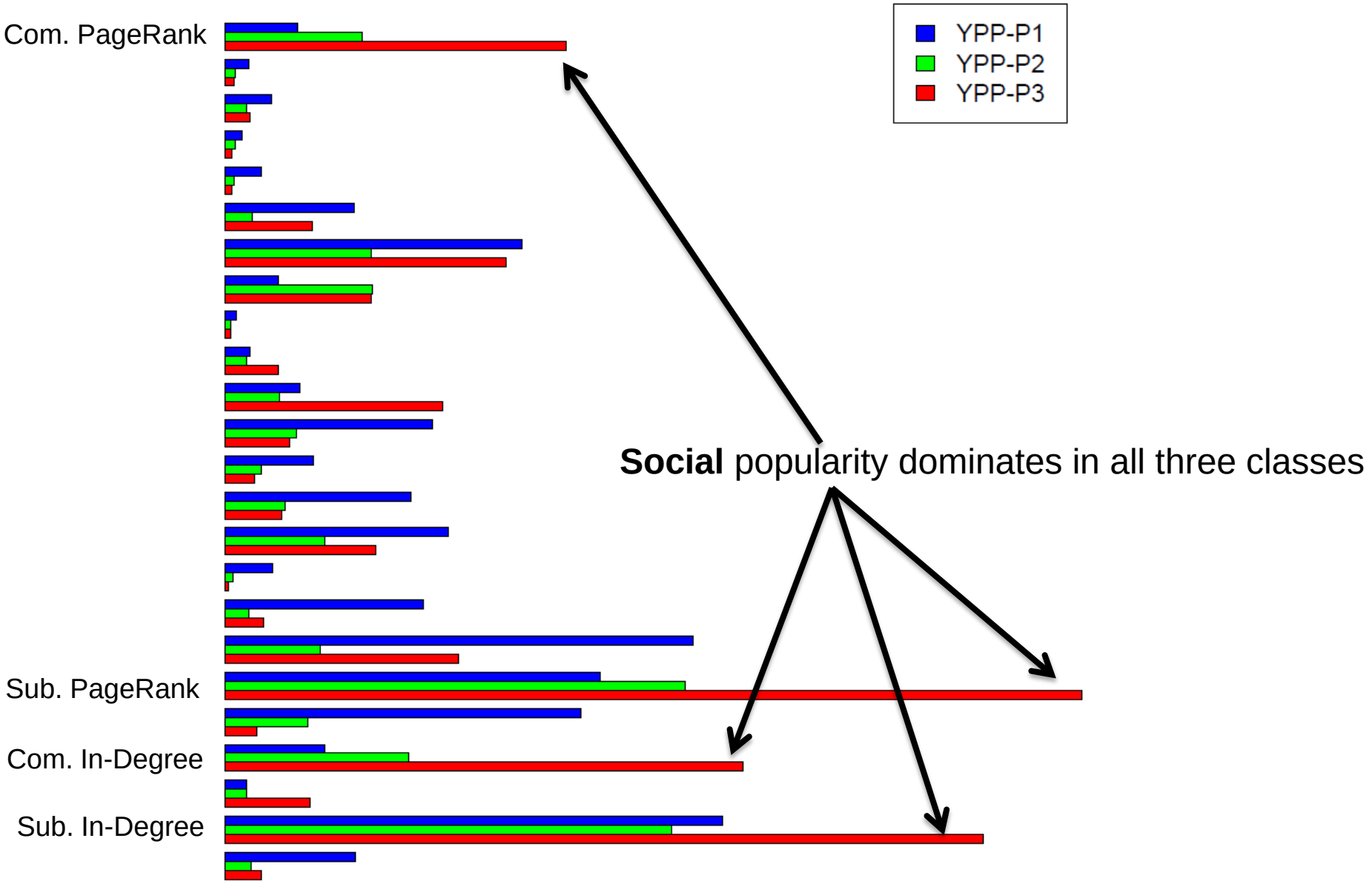
10-fold C.V. on imbalanced data

	Mean AUC	**Std. Dev.**
P1 Partners	0.968	0.0212
P2 Partner	0.958	0.0323
P3 Partners	0.943	0.0413

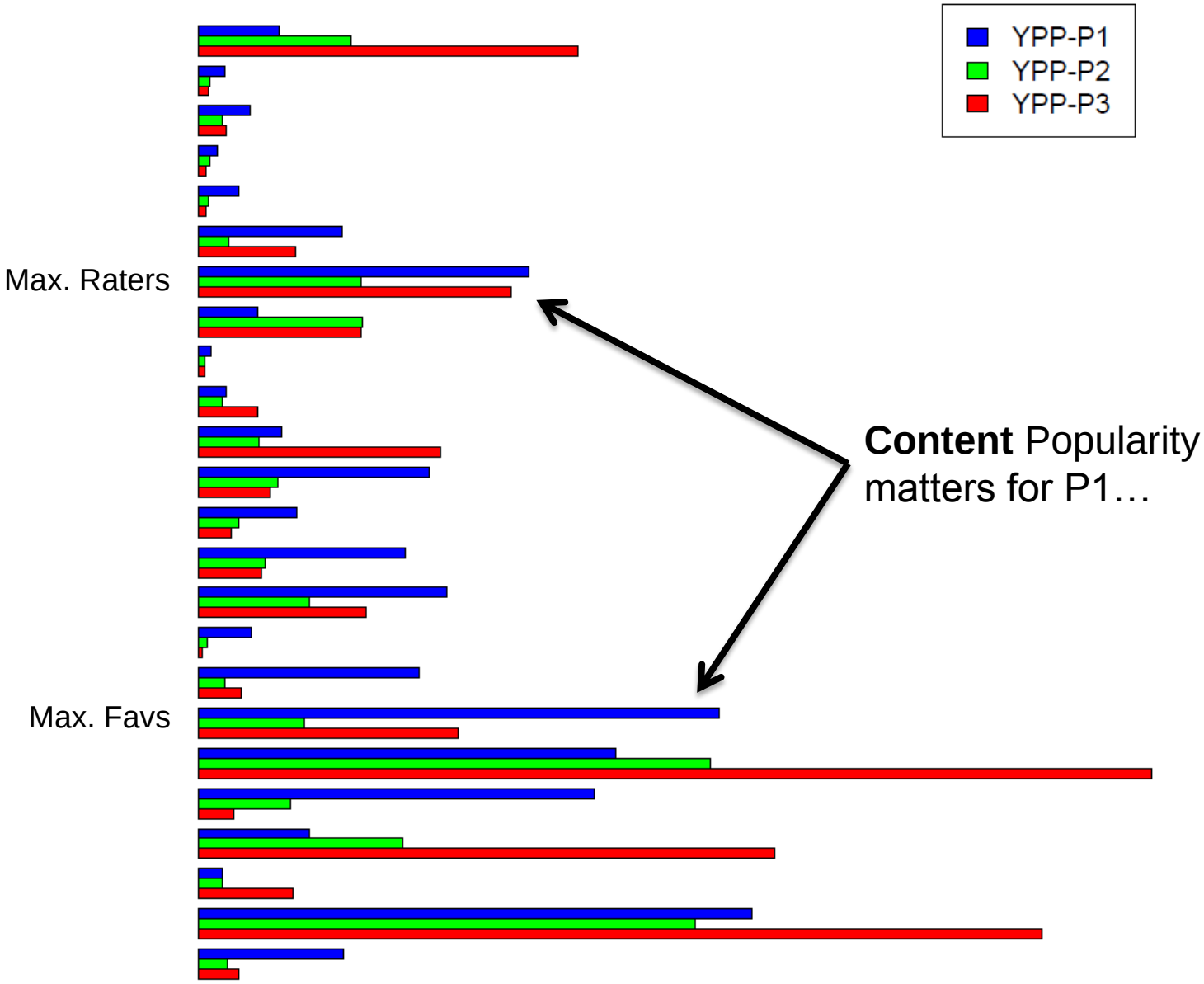
Entropy Reduction of Various Features



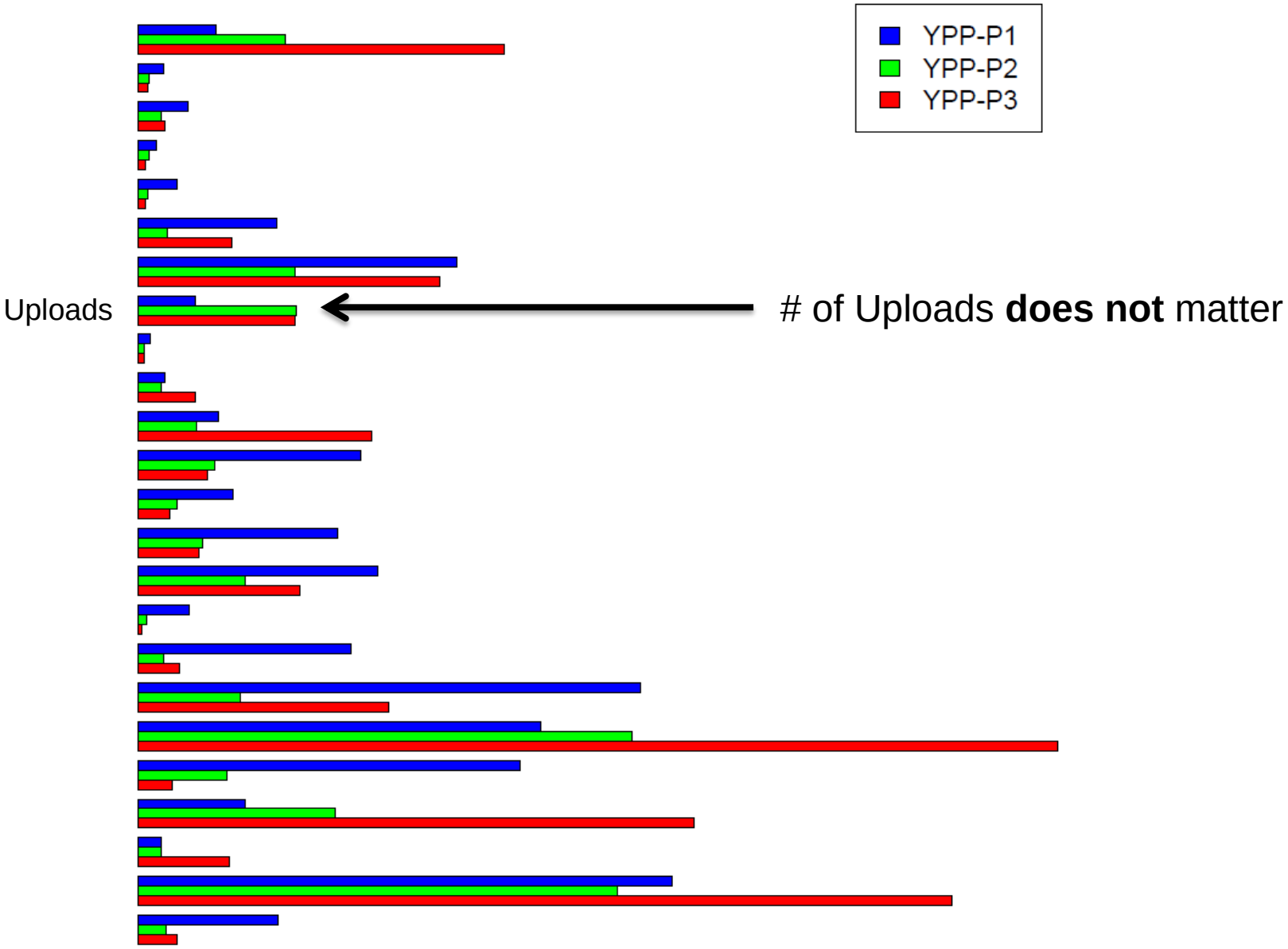
Entropy Reduction of Various Features



Entropy Reduction of Various Features



Entropy Reduction of Various Features



Take-Aways

- We found a significant **dichotomy** of social interaction and linking
- The YouTube social network **differs** from marked trait of traditional social networks - social relationships are driven by **information relationships**
- Social popularity is more in line with **top** content popularity
- **Social popularity** dominates the selection of YouTube Partners

Thanks!

Questions?

Zack Zhu

Wearable Computing Lab, ETH Zurich

zack.zhu@ife.ee.ethz.ch

http://www.ife.ee.ethz.ch/people/zazhu