

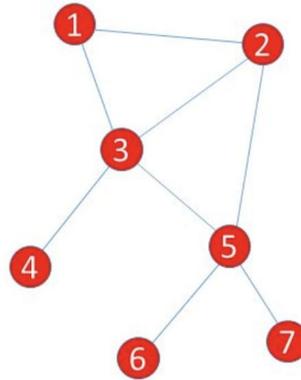
Evolution of Ego-networks in Social Media with Link Recommendations

Carlos Álvarez

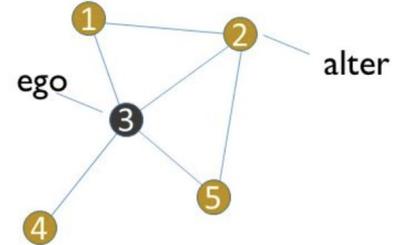
Las Ego-networks mapean la interacciones que ocurren entre contactos de un individuo



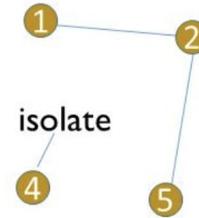
'whole' network



3's ego network



3's ego network without ego**



Permiten entender cómo está relacionado el comportamiento individual con las dinámicas sociales y grupales

El trabajo previo está relacionado sobre dinámicas globales de la red

→ Estructura y dinámica de redes sociales

- ◆ Creación de nuevos links es impulsado por reciprocidad
- ◆ Triangle closure
- ◆ Homofilia
- ◆ Métodos precisos para predicción de conexiones y recomendaciones

→ Ego-networks

- ◆ No hay muchos estudios sobre cómo se construyen redes sociales *locales*

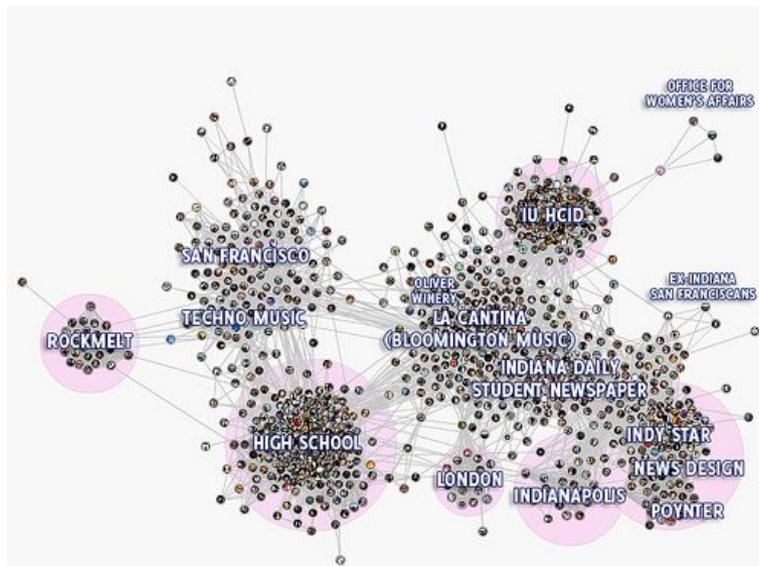
→ Efectos de sistemas recomendadores sociales

- ◆ Algoritmos basados en proximidad vs Algoritmos basados en similaridad
- ◆ Tienen un efecto sustancial sobre el crecimiento de la red.
- ◆ Nodos populares son lo que más se benefician
- ◆ Sistemas recomendadores no afectan la libertad del individuo

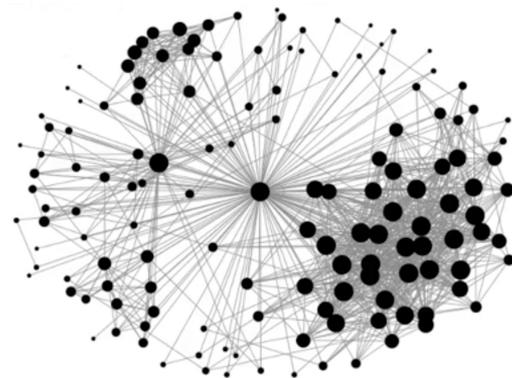
¿Cómo afectan los sistemas recomendadores a las dinámicas de las Ego-networks?

Dataset

- Tumblr: 7 Billion follow links, 130 Million public blogs
- Flickr: +500 Million links, 40 Million public profiles
- 1 blog = 1 nodo
- Acceso a al tipo de creación de un link: Espontáneo o por Recomendación



Conceptos previos



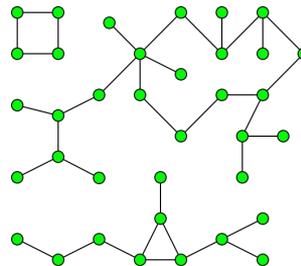
→ Teoría de grafos Diámetro: Largo máximo del camino más corto

→

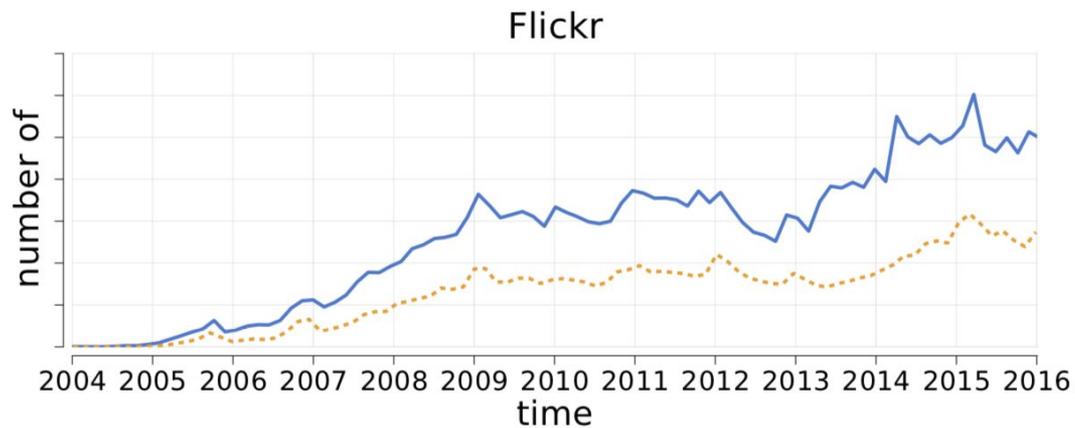
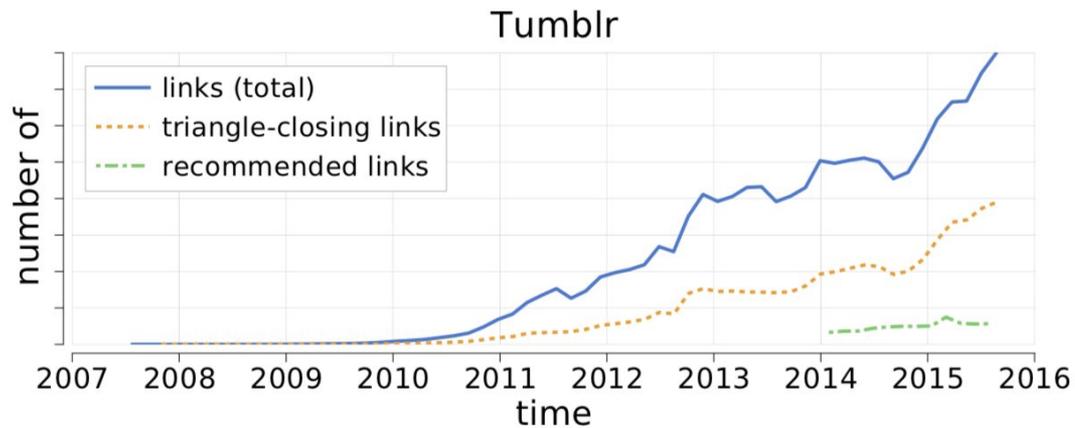
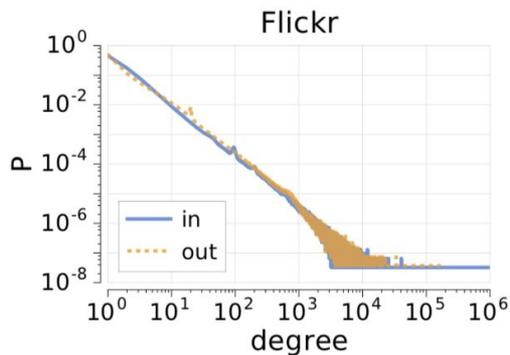
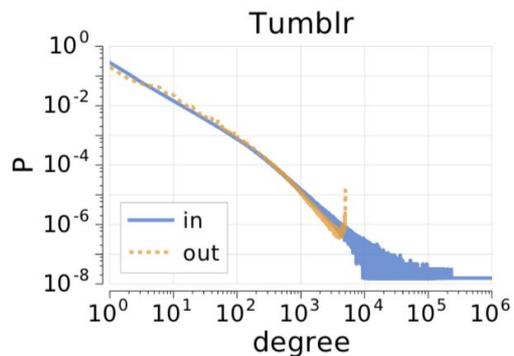
- ◆ Degree: Número de aristas incidentes al nodo.
- ◆ Distancia de la red: Distancia promedio entre todos los pares de nodos
- ◆ Densidad: (número de edges) / (máximo de edges posibles)
- ◆ Componente: maximal connected subgraph

→ Ego-network: $G_i = (N_i, E_i)$, $N_i = \Gamma_{out}(i)$, $E_i = \{(j, l) \in E | j \in \Gamma_{out}(i) \wedge l \in \Gamma_{out}(i)\}$

→ Distinción de links espontáneos y recomendados



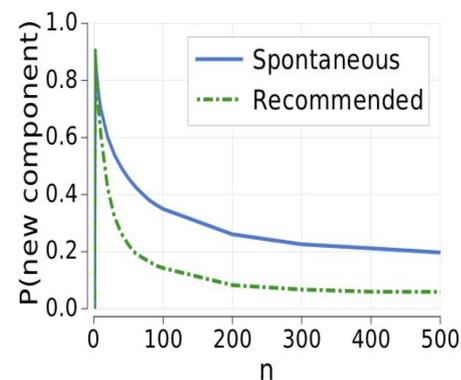
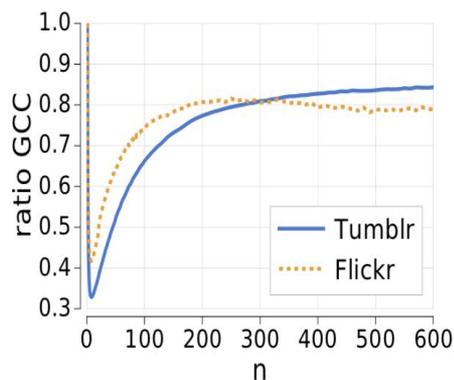
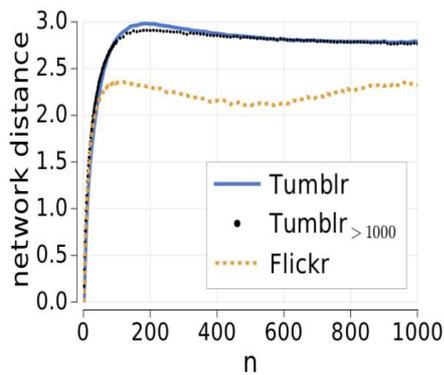
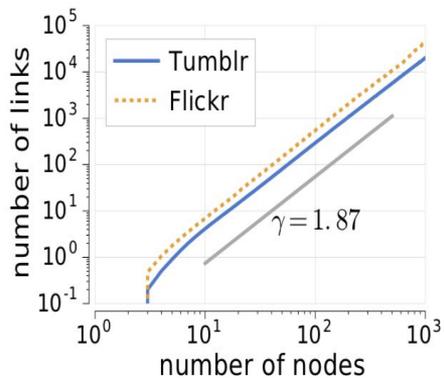
Data Overview



Interrogantes

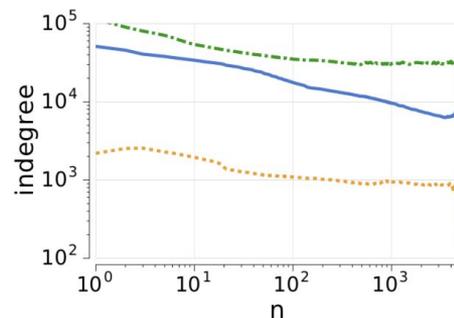
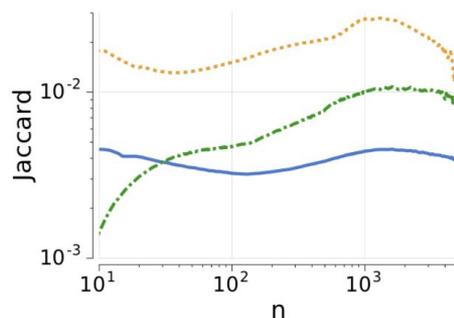
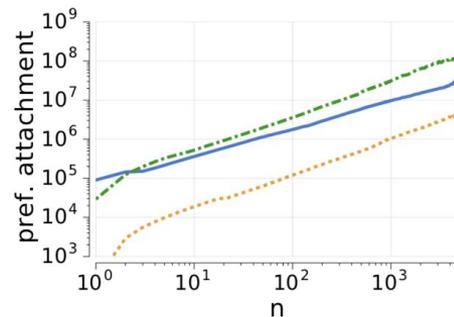
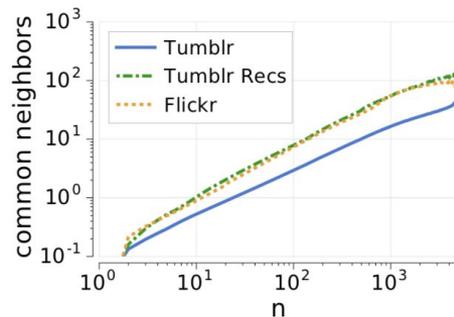
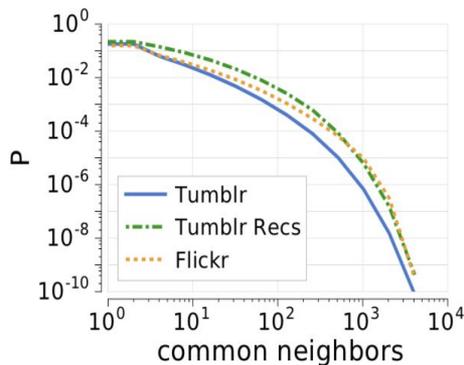
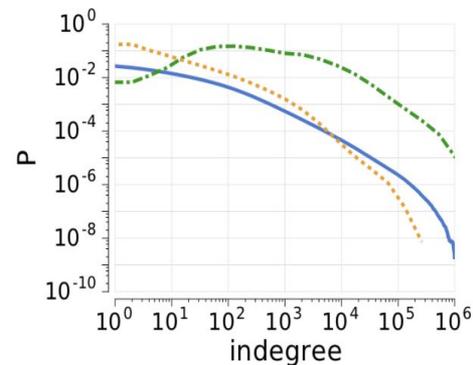
1. ¿Cómo evoluciona la densidad, diámetro y estructura de componentes a medida que la Ego-network crece?
2. ¿Cómo cambia el criterio de elección de vecinos a medida que que la Ego-network crece?
3. ¿Cuándo se expanden las Ego-networks?
4. ¿El crecimiento de las Ego-networks es impulsada por los límites de sus comunidades?
5. ¿Los links recomendados fomentan la diversidad?

1. ¿Cómo evoluciona la densidad, diámetro y estructura de componentes a medida que la Ego-network crece?



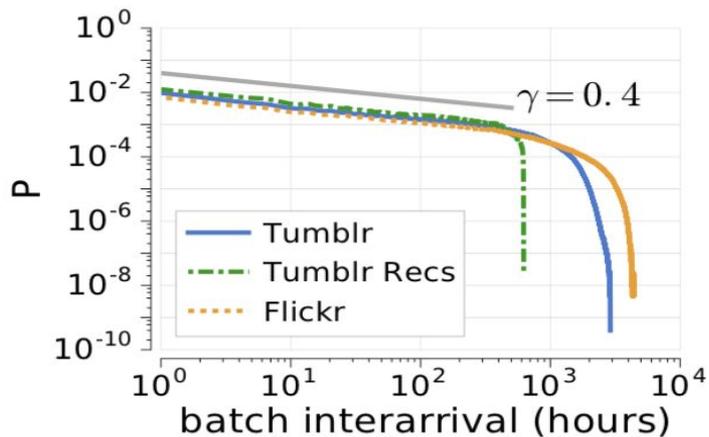
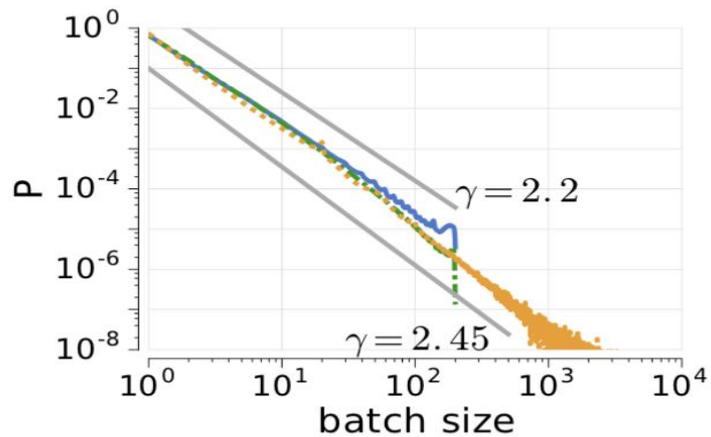
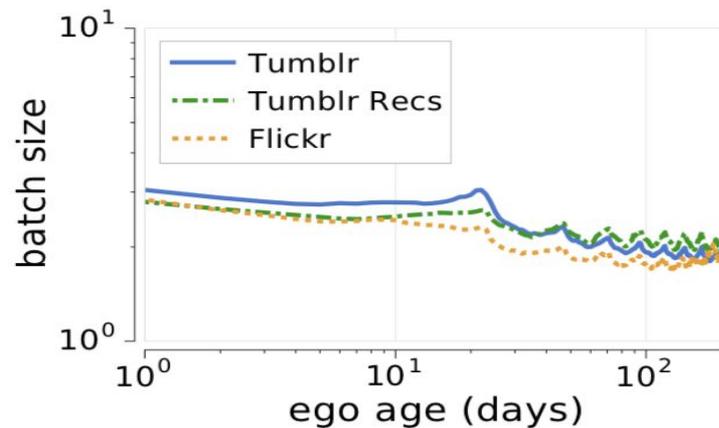
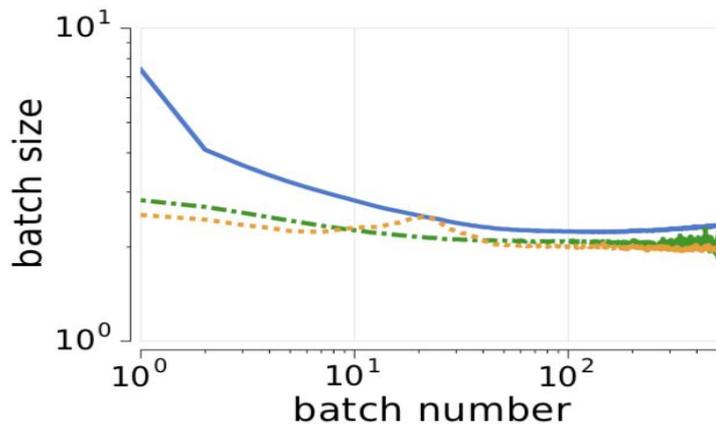
La evolución de la ego-network se experimenta 3 fases:
Exploración, consolidación, estabilidad

2. ¿Cómo cambia el criterio de elección de vecinos a medida que que la Ego-network crece?

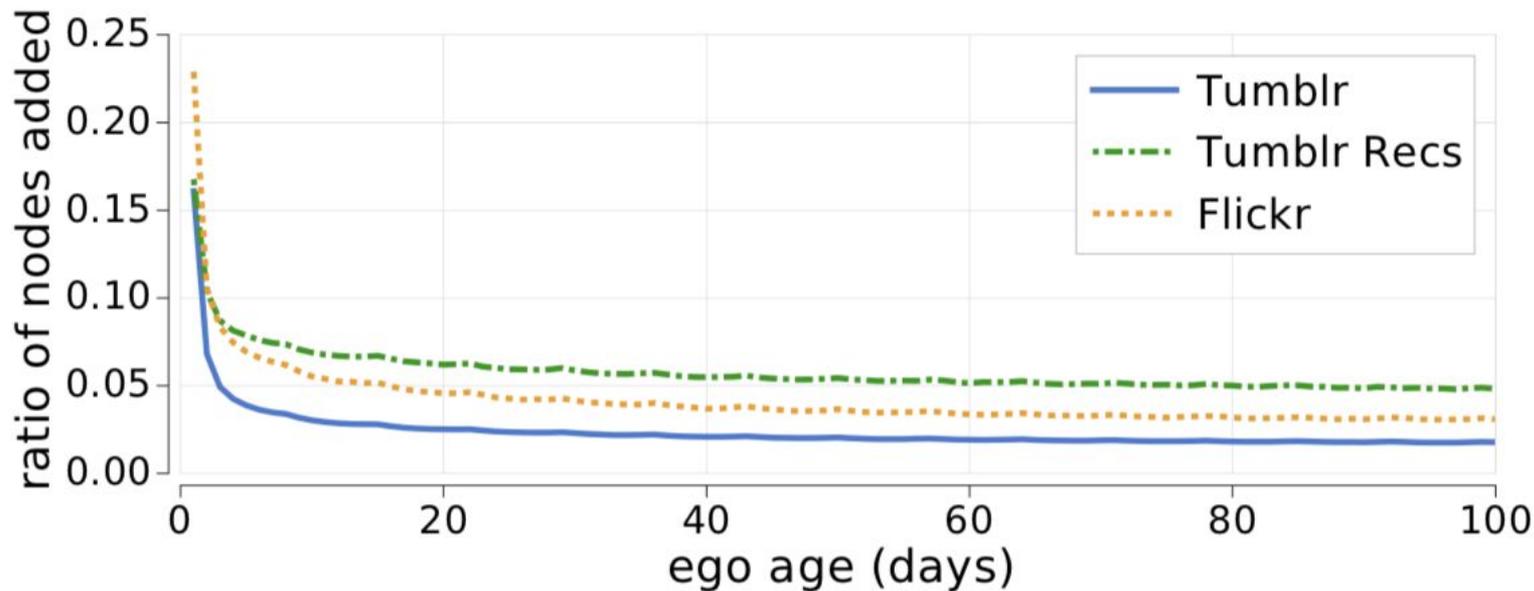


“Rich get richer”: Los links creados por recomendación tienden a beneficiar a los nodos más populares

3. ¿Cuándo se expanden las Ego-networks?

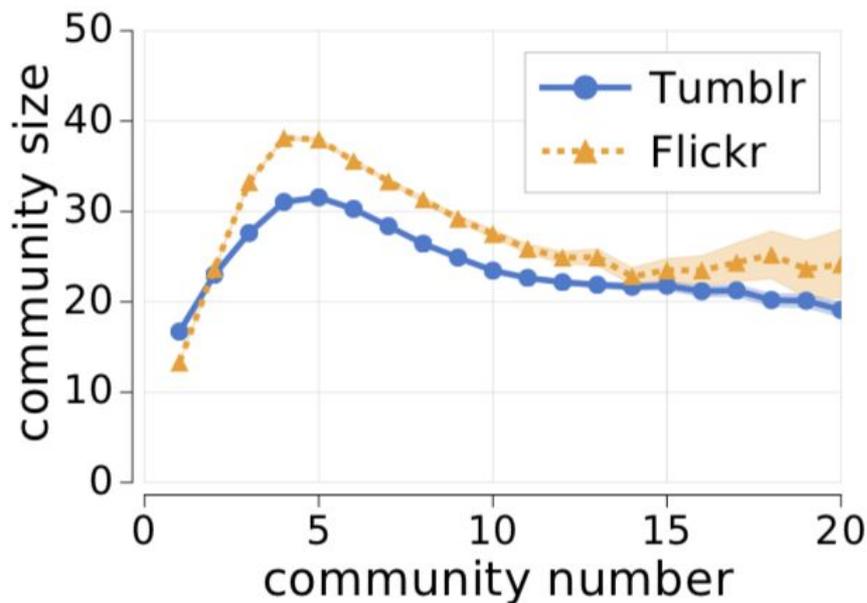
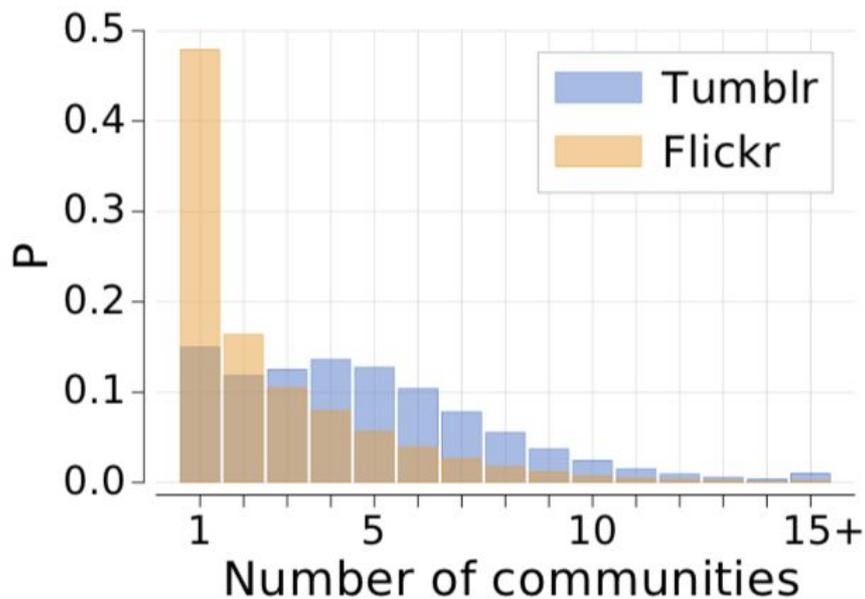


3. ¿Cuándo se expanden las Ego-networks?

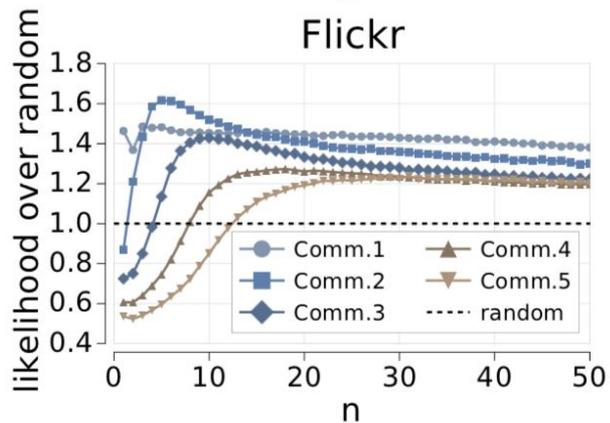
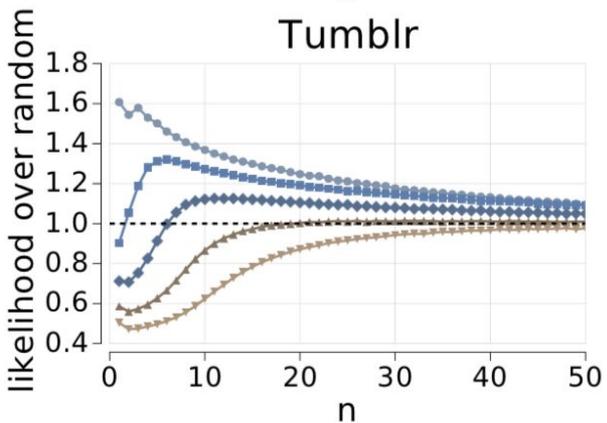
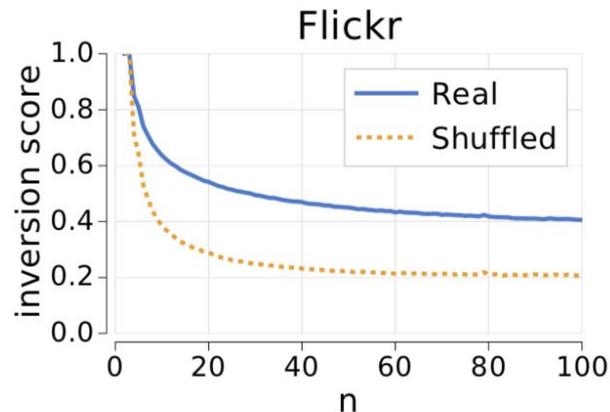
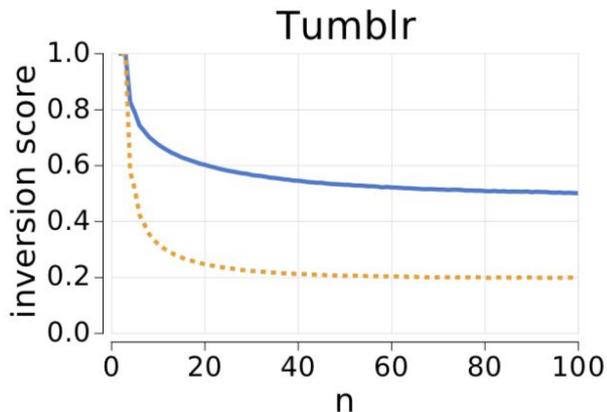


Grandes cantidades de links se crean en los primeros días, luego tiende a estabilizarse uniformemente

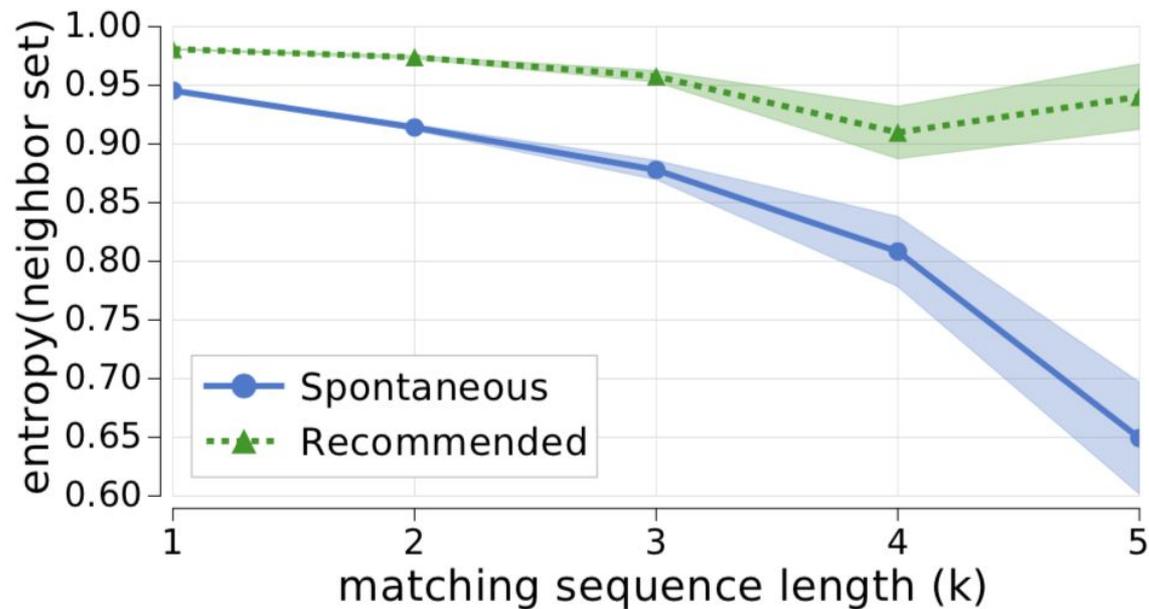
4. ¿El crecimiento de las Ego-networks es impulsada por los límites de sus comunidades?



4. ¿El crecimiento de las Ego-networks es impulsada por los límites de sus comunidades?



5. ¿Los links recomendados fomentan la diversidad?



Sis. recs. exponen a los usuarios a un mayor set de potenciales conexiones

Impacto en las predicciones

1. ¿Hasta qué punto las características temporales incrementan la habilidad de predecir nuevas conexiones?
2. ¿Es posible limitar el sesgo del sistema recomendador, mientras se mantiene su alta precisión?

1. ¿Hasta qué punto las características temporales incrementan la habilidad de predecir nuevas conexiones?

Model	AUC	F-Score
Baseline	0.893	0.813
+ age	0.938	0.864
+ k_{out}	0.897	0.817
All	0.943	0.87

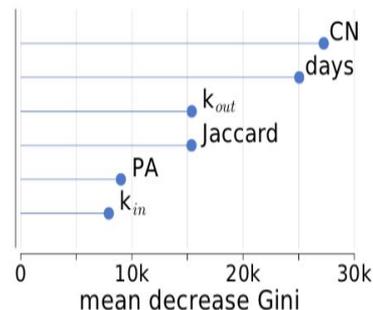


Table 1: Link prediction results.

Figure 15: Feature importance.

Incorporando variables de tiempo al sistema predictor de futuras conexiones, este mejora su rendimiento en un 5.6% y 7% en los indicadores de AUC y F-Score sobre el modelo base

2. ¿Es posible limitar el sesgo del sistema recomendador, mientras se mantiene su alta precisión?

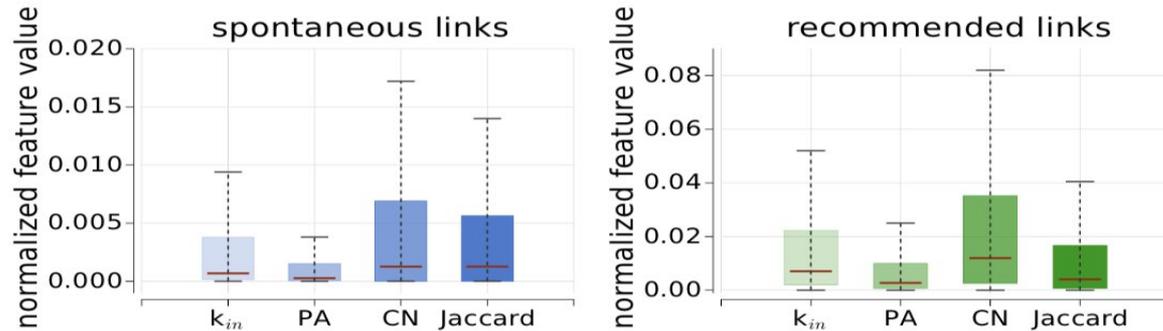


Figure 16: Distribution of structural features of spontaneous and recommended links. Given directed links in the form (i, j) we show the boxplots of the distributions of $k_{in}(j)$, $PA(i, j)$, $CN(i, j)$, and $Jaccard(i, j)$.

Añadir variables temporales y estructurales permitirán generar algoritmos con menor sesgo, cuyas sugerencias simulen mejor el proceso de link espontáneos

Conclusiones

- Se expuso un análisis de las dinámicas de la evolución en ráfagas, impulsadas por la comunidad, expansión de diámetro y selección de nuevos nodos en una escala de tiempo.
- Links provenientes de recomendaciones tienen diferentes propiedades estadísticas comparados con los generados espontáneamente.
- Links provenientes de recomendación fomentan la amplitud y diversidad de la red
- Limitaciones:
 - ◆ Flickr y Tumblr son redes basadas en contenido de gustos, a diferencia redes de perfiles únicos donde la gente se conoce en la vida real. Ej. Facebook, LinkedIn.
 - ◆ El nodo no incorpora información del perfil.
- Implicaciones:
 - ◆ Modelos de crecimiento de redes sociales deberían incluir pesos diferentes a links generados espontáneamente vs algorítmicamente.
 - ◆ Se provee una intuición sobre cómo los sistemas recomendadores de links podrían incorporar señales de la evolución de las ego-networks para mejorar la calidad de sus predicciones.

Evolution of Ego-networks in Social Media with Link Recommendations

Carlos Álvarez