Evaluación de Recomendadores Centrada en el Usuario

Denis Parra
IIC 3633, Sistemas Recomendadores
PUC Chile
2016

Agenda Semestral

Week	Fecha semana	Clase Martes	Clase Jueves	Presentador 1	Presentador 2	Presentador 3
I	2 - 4 Ago	Intro + CF	CF + Clustering			
II	9 - 11 Ago	CF item-based	Slope One + RecSys			
III	16 - 18 Ago	Evaluacion de RecSys	Evaluacion de RecSys			
IV	23 - 25 Ago	Content-based	Tag-based			
V	30 Ag - 1 Sept	Hybrid	Factorizacion Matricial			
VI	6 - 8 Sept	Context-aware RecSys	Implicit Feedback			
VII	13 - 15 Sept	student presentation (Context, MF	RECSYS Conf	V. Dominguez	J. Schellman	P. Lopez
VIII	20 - 22 Sept	RECSYS Conf	student presentation (IF, MF)	~~F. Lucchini~~	~~V. Claro~~	~~V. Castillo~~
IX	27 - 29 Sept	Presentaciones: Proy. Final	Presentaciones: Proy. Final			
X	4 - 6 Oct	User-centric RecSys/Interfaces	student presentation	J. Lee	C. Kutscher	~~R. Carmona~~
XI	11 - 13 Oct	Active Learning/Ranking	student presentation	F. Rojos	J. Navarro	N. Morales
XII	18 - 20 Oct	Graph-based	student presentation	P. Messina	S. Martí	J. Castro
XIII	25 - 27 Oct	Applications: Social/Trust/Music	student presentation	J.M. Herrera	V. Dragicevic	L. Zorich
XIV	1 - 3 Nov	Applications: POI/Tourism	student presentation	I. Becker	T. Hepner	M. Troncoso
XV	8 - 10 Nov	Applications: Educ/Soft.Eng.	student presentation	R. Perez	P. Sanabria	J. Diaz
XVI	15 - 17 Nov	Deep Learning	student presentation	Felipe del Río	L. Pose	G. Sepulveda
XVII	29 Nov - 1 Dic	Presentacion Final	Presentacion Final			

Temas

- Transparencia y Explicabilidad
- Controlabilidad
- Visualizaciones e Interactividad
- Algunos ejemplos para evaluación de la experiencia del usuario
- Frameworks para evaluación
  - Pearl Pu
  - Bart Knijnenburg

Por qué evaluación centrada en el usuario?

- Mayoría de investigación evalúa resultado de recomendaciones off-line.
- Mejoras pequeñas de predicción en los algoritmos no siempre se traducen en una mejor percepción de los usuarios (Konstan & Riedl 2012)
- La precisión de los algoritmos es sólo uno de los factores que influencian la aceptación de las recomendaciones por parte de los usuarios

Explicabilidad

- Capítulo en "HandBook of Recommender Systems" [Tintarev & Masthoff, 2012]
- Ellas proponen algunas direcciones generales para diseñar explicaciones para SisRec
  - Considerar beneficios a obtener (propósito)
  - Evitar (o buscar) relación con funcionamiento del recomendador
  - Presentación y forma de interacción
  - Relación entre algoritmo y tipo de explicaciones

1. Criterios de Explicación

Propósito	Descripción
1.1 Transparencia	Explicar cómo funciona el sistema
1.2 Escrutabilidad	Dejar al usuario indicar que el sistema comete un error
1.3 Confianza	Incrementar confianza del usuario en el sistema
1.4 Efectividad	Ayudar al usuario a tomar buenas decisiones
1.5 Persuasión	Convencer a usuario a probar o a comprar
1.6 Eficiencia	Ayudar a usuarios a tomar decisiones más rápido
1.7 Satisfacción	Aumentar facilidad de uso o el disfrute en el sistema

1.1 Transparencia

- Ejemplo a partir de artículo del Wall Street Journal:

"*If TiVo Thinks You Are Gay, Here's How to Set It Straight*"

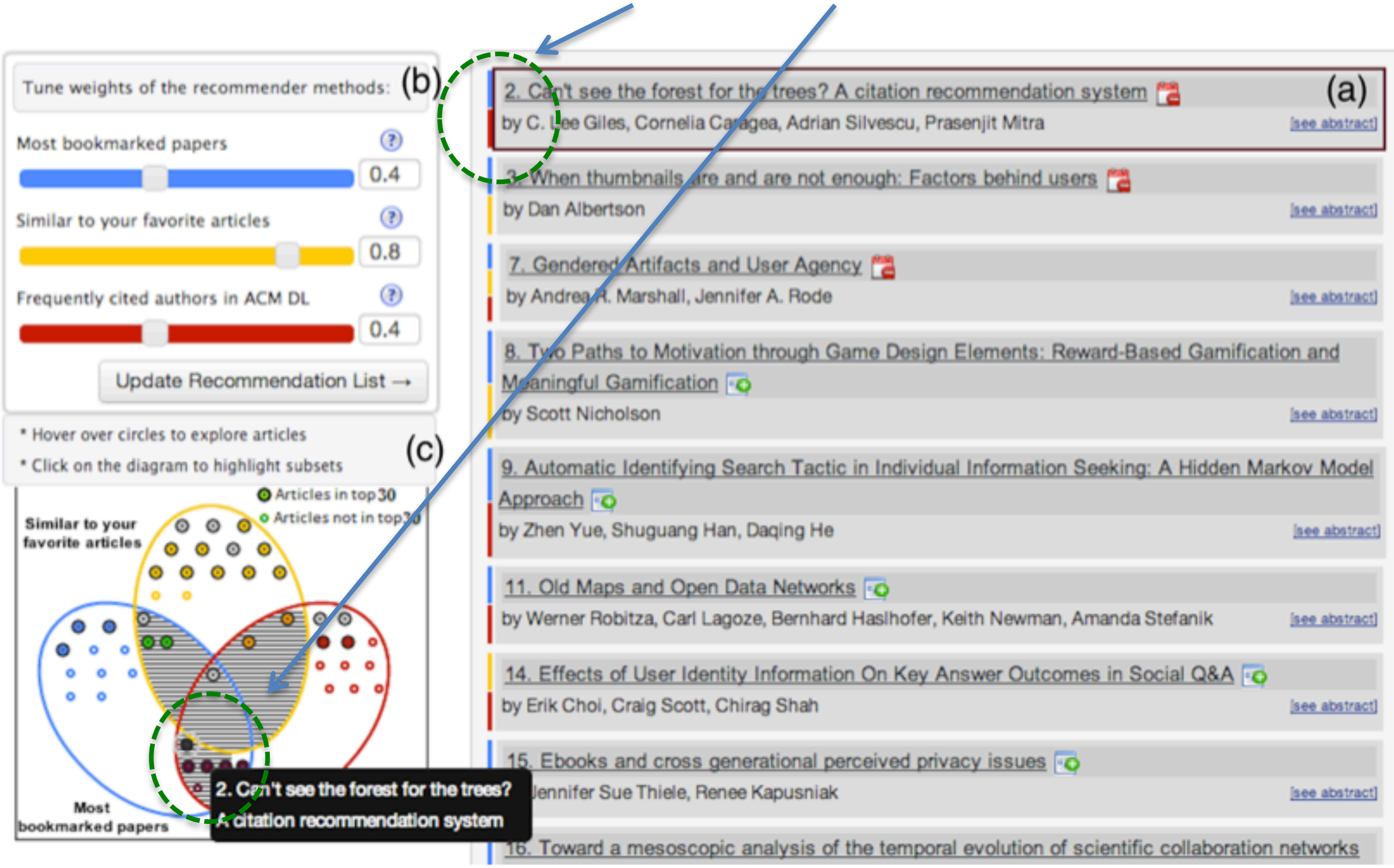
- Un usuario sospechó que TiVo pensó que él era homosexual pues el sistema comenzó a grabar automáticamente estos programas.
- En el artículo, se explica que este es un caso en que un usuario podría requerir transparencia en el algoritmo recomendador.

Escrutabilidad

- Permitir al usuario inpeccionar o “escrutar” el resultado de la recomendación
- Si bien está relacionado con transparencia, se sugiere identificar y separarlo como ítem.

Fig. 15.1: Scrutable holiday recommender [21]. The explanation is in the circled area, and the user profile can be accessed via the “why” links.

Escrutabilidad



SetFusion: A Controllable Hybrid Recommender

Parra, D., Brusilovsky, P., Trattner, C.

IUI 2014, Haifa, Israel

Confianza

- Mayor transparencia y posibilidad de interactuar con el recomendador está asociado en varios estudios con mayor confianza en el sistema
- Podría estar asociado directamente a la precisión de predicción de la recomendación, pero no siempre!
- Una buena métrica de confianza: Lealtad del usuario en volver a usar el sistema

Confianza

- Dos trabajos muestran que confianza/satisfacción y predicción no siempre están correlacionados

McNee et al. **Don't look stupid: avoiding pitfalls when recommending research papers**. CSCW (2006)

Cramer et al. **The effects of transparency on trust in and acceptance of a content-based art recommender**. UMUAI 18(5), 455–496 (2008).

Persuasión

- Uno de los primeros trabajos en el área de “explicabilidad” de recomendaciones intentaba explicar al usuario las recomendaciones hechas; probaron 21 métodos posibles.
- El autor del paper en algún momento llamó la atención de no considerar ese estudio como el modelo de explicabilidad, ya que hacer al usuario consciente de una decisión y persuadirlo puede tener efectos importantes

Persuasión II

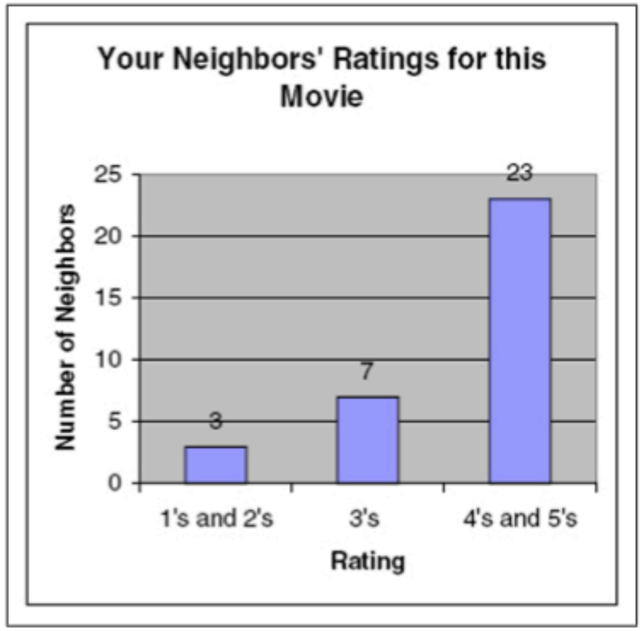


Fig. 15.2: One out of twenty-one interfaces evaluated for persuasiveness - a histogram summarizing the ratings of similar users (neighbors) for the recommended item grouped by good (5's and 4's), neutral (3's), and bad (2's and 1's), on a scale from 1 to 5 [29].

Herlocker,J.L.,Konstan,J.A.,Riedl,J.:**Explaining collaborative filtering recommendations**.

In: ACM conference on Computer supported cooperative work, pp. 241–250 (2000)

Efectividad

- Conectado con la definición anterior, la explicación/persuasión de una recomendación debiese estar asociada a una buena percepción del usuario
- “Vig et al. measure perceived effectiveness: “This explanation helps me determine how well I will like this movie.” [62]. ”
- Se podría medir como la diferencia entre la percepción del ítem al momento de elegirlo y después del consumo.

Efectividad II

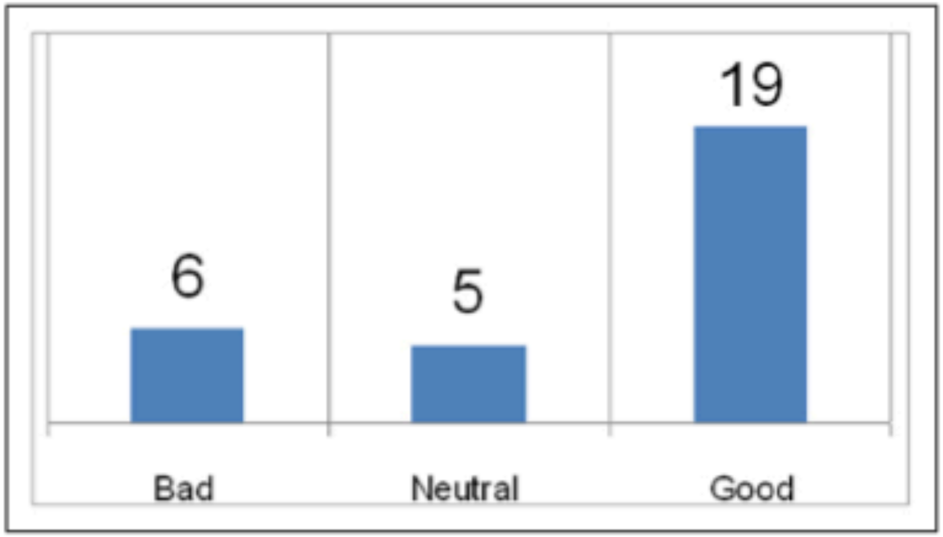


Fig. 15.3: The Neighbor Style Explanation - a histogram summarizing the ratings of similar users (neighbors) for the recommended item grouped by good (5's and 4's), neutral (3's), and bad (2's and 1's), on a scale from 1 to 5. The similarity to Figure 15.2 in this study was intentional, and was used to highlight the difference between persuasive and effective explanations [11].

Efectividad III

Table 15.3: The keyword style explanation by [11]. This recommendation is explained in terms of keywords that were used in the description of the item, and that have previously been associated with highly rated items. "Count" identifies the number of times the keyword occurs in the item's description, and "strength" identifies how influential this keyword is for predicting liking of an item.

Word	Count	Strength	Explain
HEART	2	96.14	*Explain*
BEAUTIFUL	1	17.07	*Explain*
MOTHER	3	11.55	*Explain*
READ	14	10.63	*Explain*
STORY	16	9.12	*Explain*

Title	Author	Rating	Count
Hunchback of Notre Dame	Victor Hugo, Walter J. Cobb	10	11
Till We Have Faces: A Myth Retold	C.S. Lewis, Fritz Eichenberg	10	10
The Picture of Dorian Gray	Oscar Wilde, Isobel Murray	8	5

Eficiencia

- Bajo este parámetro, los tipos de explicaciones debieran optimizarse por dominio para elegir entre opciones que compiten. Por ejemplo, en cámaras

<<*"Less Memory and Lower Resolution and Cheaper"* >>

Altamente usado en "Conversational" SisRec, donde el usuario refina iterativamente sus preferencias.

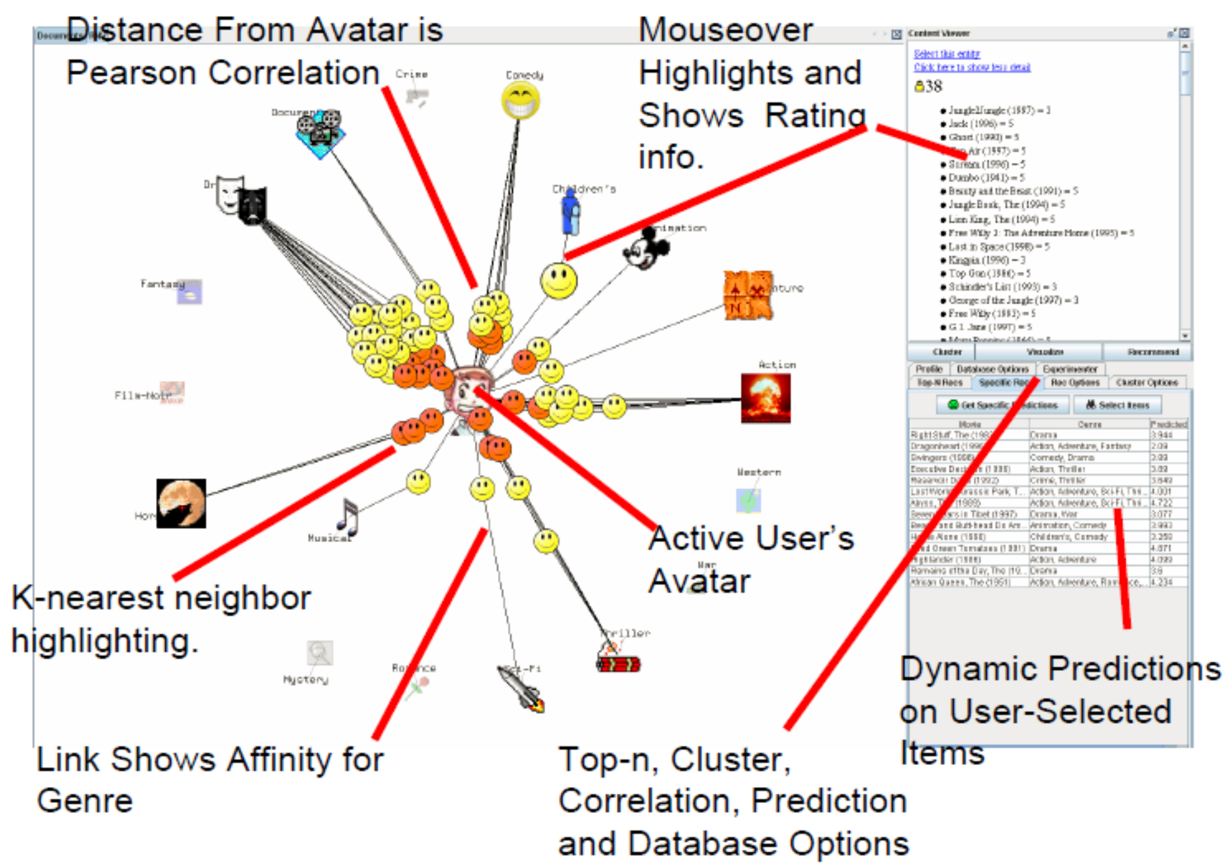
Satisfacción

- Esta es probablemente la métrica que resumen de mejor forma el objetivo de un sistema recomendador
- Existen algunos instrumentos (cuestionarios con varios sets de preguntas) que intentan medir esta dimensión. Lo veremos en más detalle en User Centric Evaluation Frameworks.

Visualizaciones

Related work on Visual RS - 1

- 2008: PeerChooser (CHI 2008)
- John O'Donovan and Barry Smyth (UCD)
- Brynjar Gretarsson, Svetlin Bostandjiev, Tobias Hollerer (UCSB)

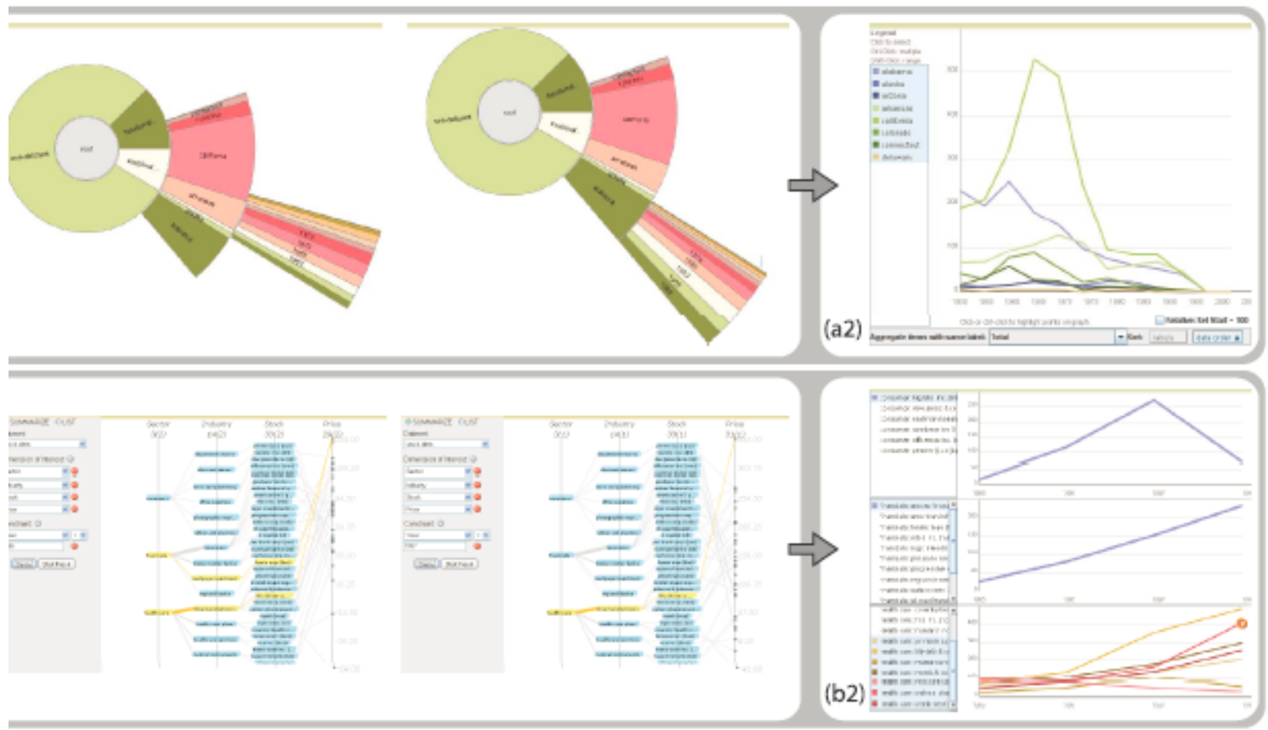


2: Annotated Screenshot of PeerChooser's Interactive Interface.

Related work on Visual RS - 2

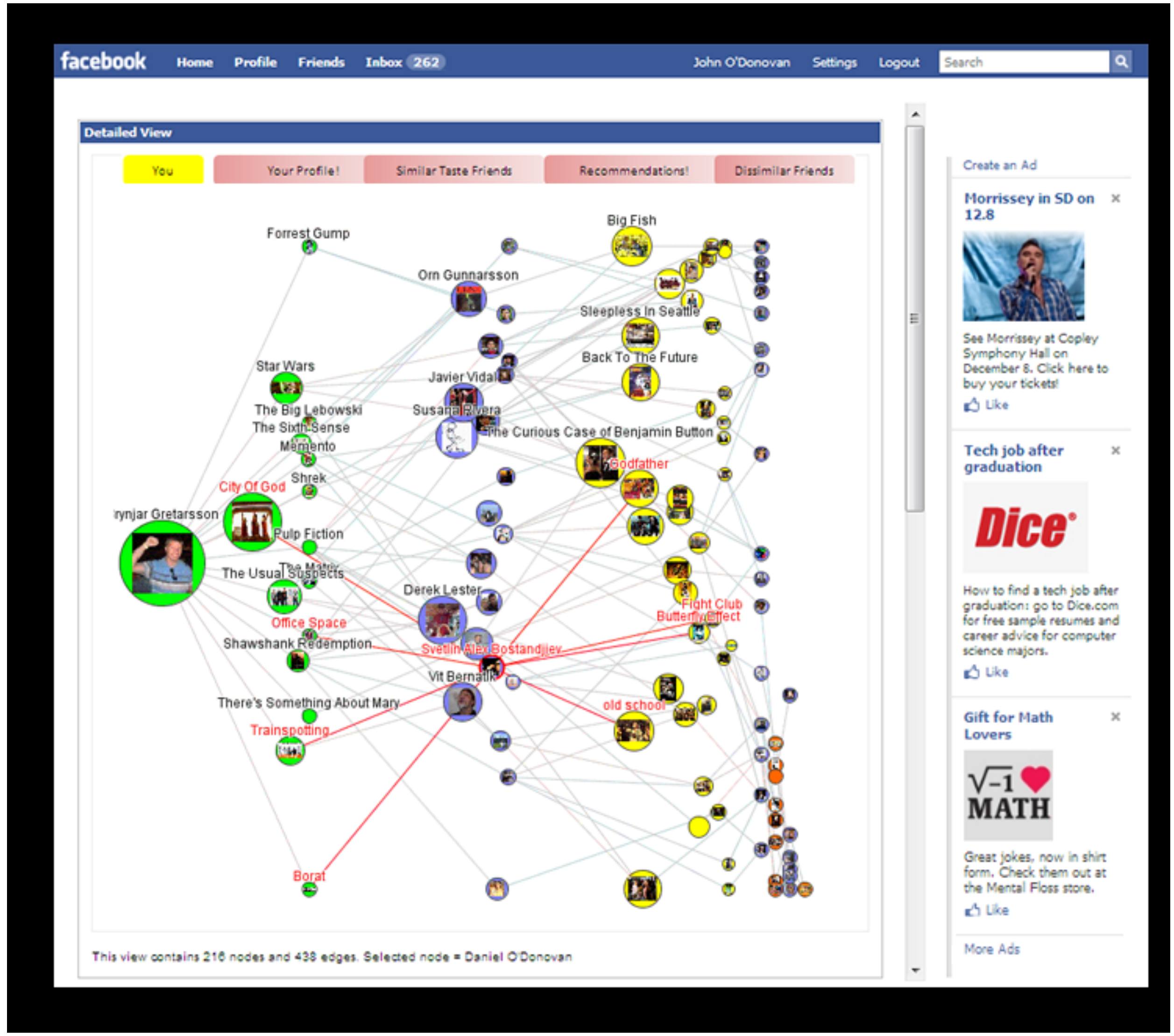
- 2009: Behavior-driven Visualization Recommendations (IUI 2009)
- David Gotz, Zhen Wen (IBM Research)

Given certain tasks inferred from user's behavior, recommend visualizations to accomplish those tasks more efficiently



Related work on Visual RS – 2

- 2010: “SmallWorlds: Visualizing Social Recommendations” IEEE-VGTC 2010
- Brynjar Gretarsson, John O'Donovan , Svetlin Bostandjiev, Christopher Hall, Tobias Höllerer(UCSB)
- User study with 17 users



Related work on Visual RS - 3

- 2010: Pharos "Who is Talking about What: Social Map-based Recommendation for Content-Centric Social Websites" (RecSys 2010)
- Zhao et al.(IBM China)

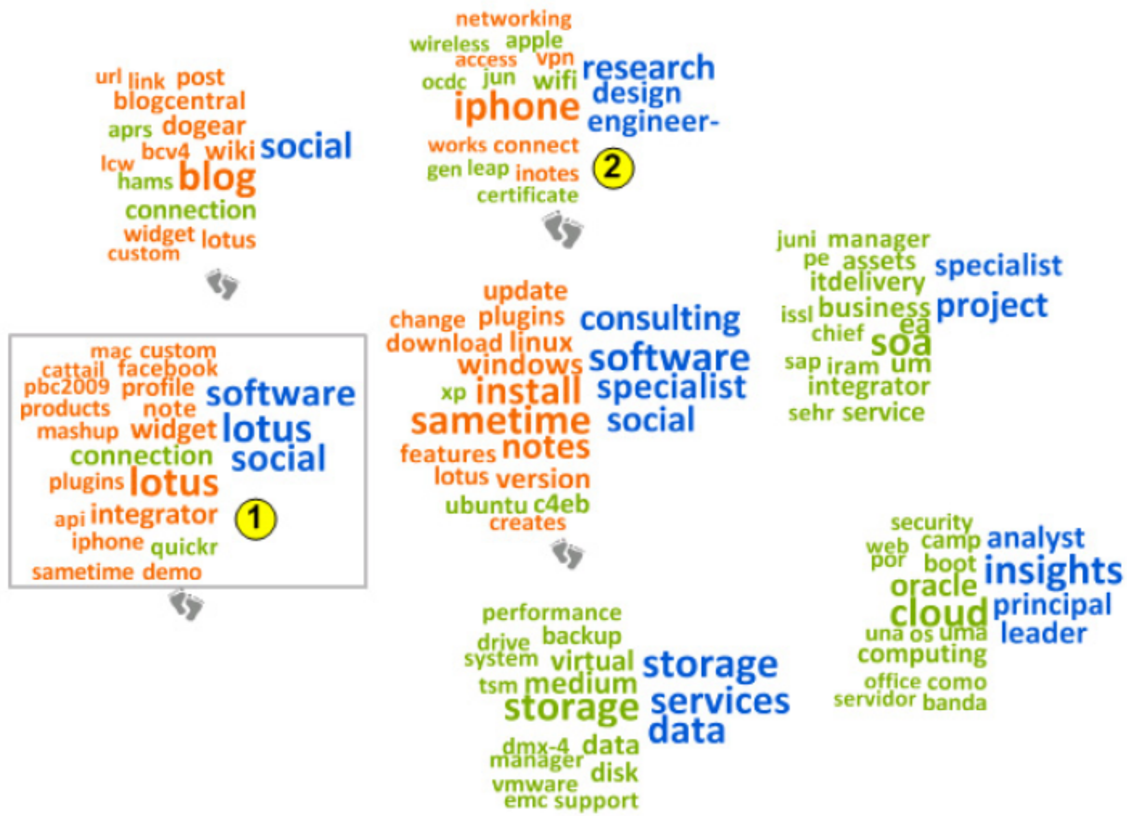


Figure 2: Highlight a user's activities (keywords in orange) in multiple communities. The size of the footprint indicates how active the user is in the attached community.

Related Work – 3.5 ☺

- **2010: Opinion Space: A Scalable Tool for Browsing Online Comments**
- Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, Ken Goldberg
- Software sponsored by US Government to diversify political opinnions

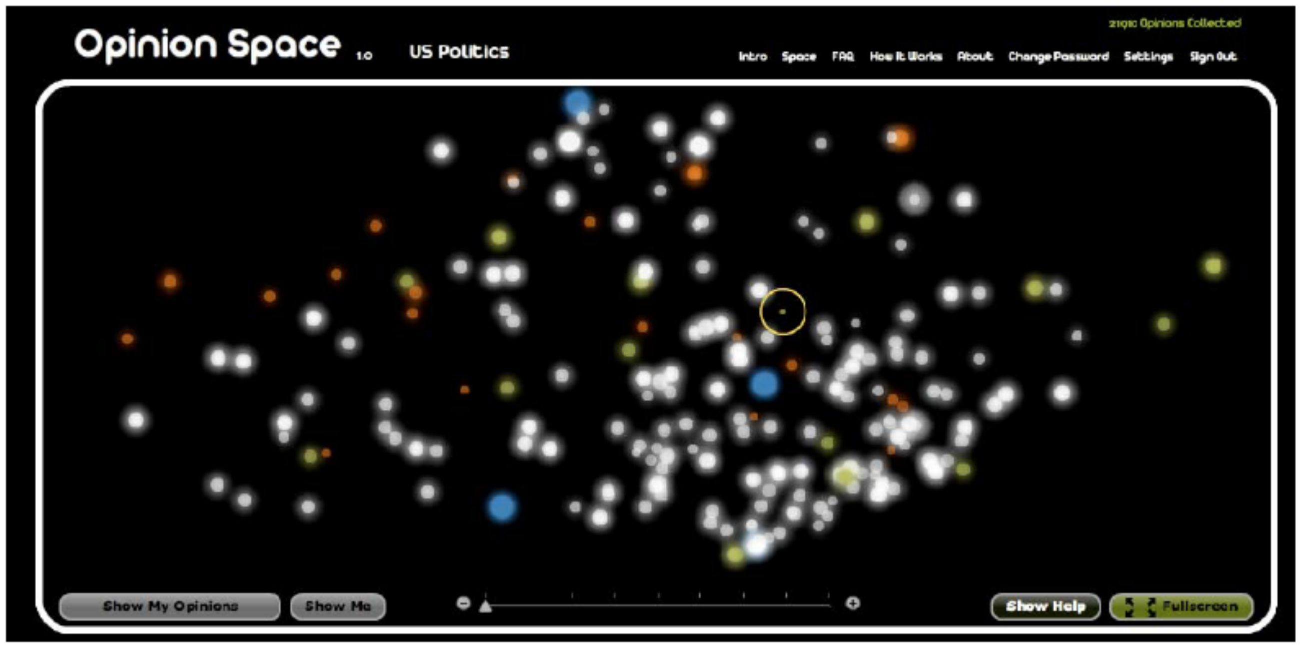
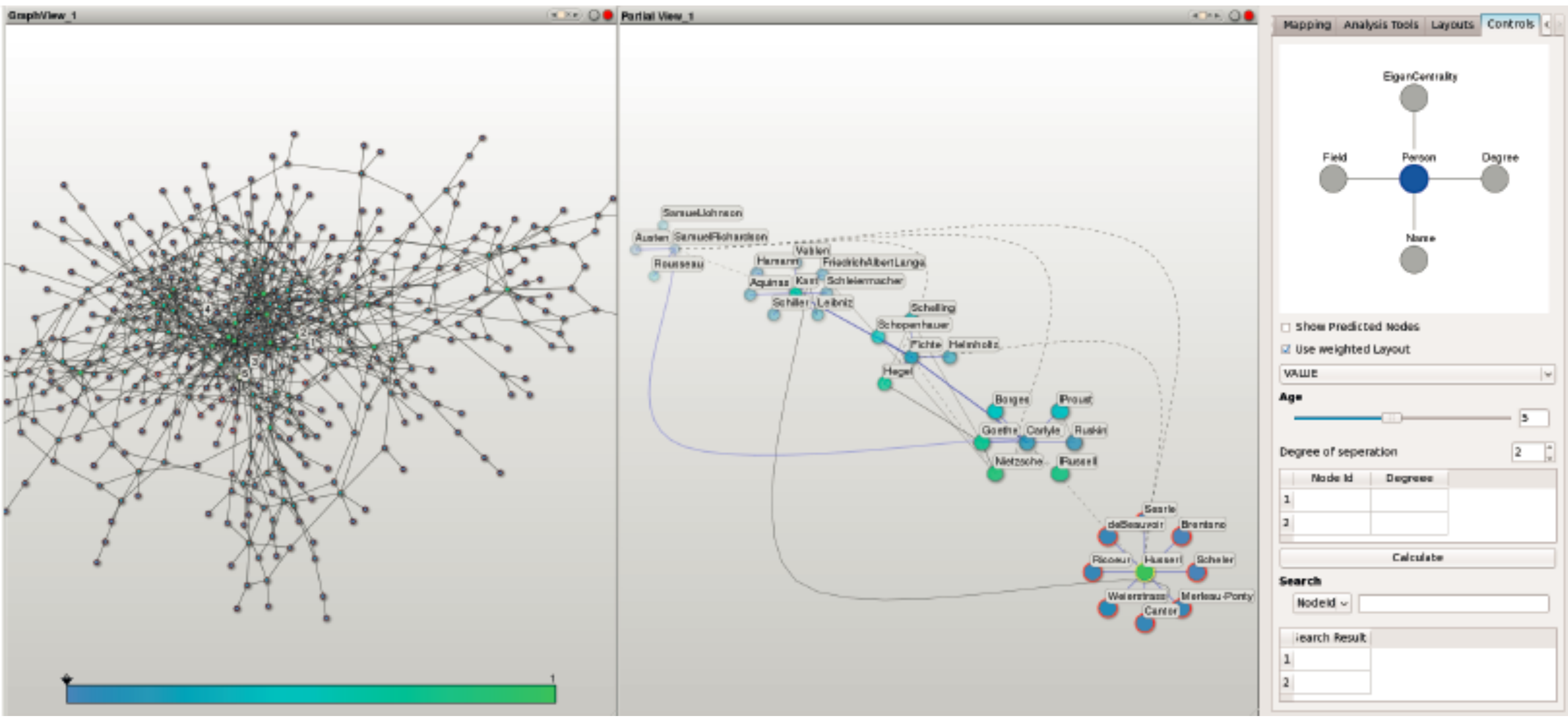


Figure 1. A screenshot of the Opinion Space 1.0 interactive map. Each point corresponds to a user and comment. The point with the halo indicates the position of the active user; green points correspond to comments rated positively by the active user, and red points correspond to comments rated negatively. Larger and brighter points are associated with the comments that are rated more positively by the user community.

Related work on Visual RS - 4

- 2011: ***Visual Recommendations for Network Navigation***. IEEE Symposium on Visualization . Tarik Crnovrsanin, Isaac Liao, Yingcai Wu, Kwan-Liu Ma
- Build on top of netzen: http://vis.cs.ucdavis.edu/~correac/netzen/index.html



Related work on Visual RS - 5

- 2011: SFViz: interest-based friends exploration and recommendation in social networks SFVIZ (VINCI 2011)
- Gou, You (?) et al.

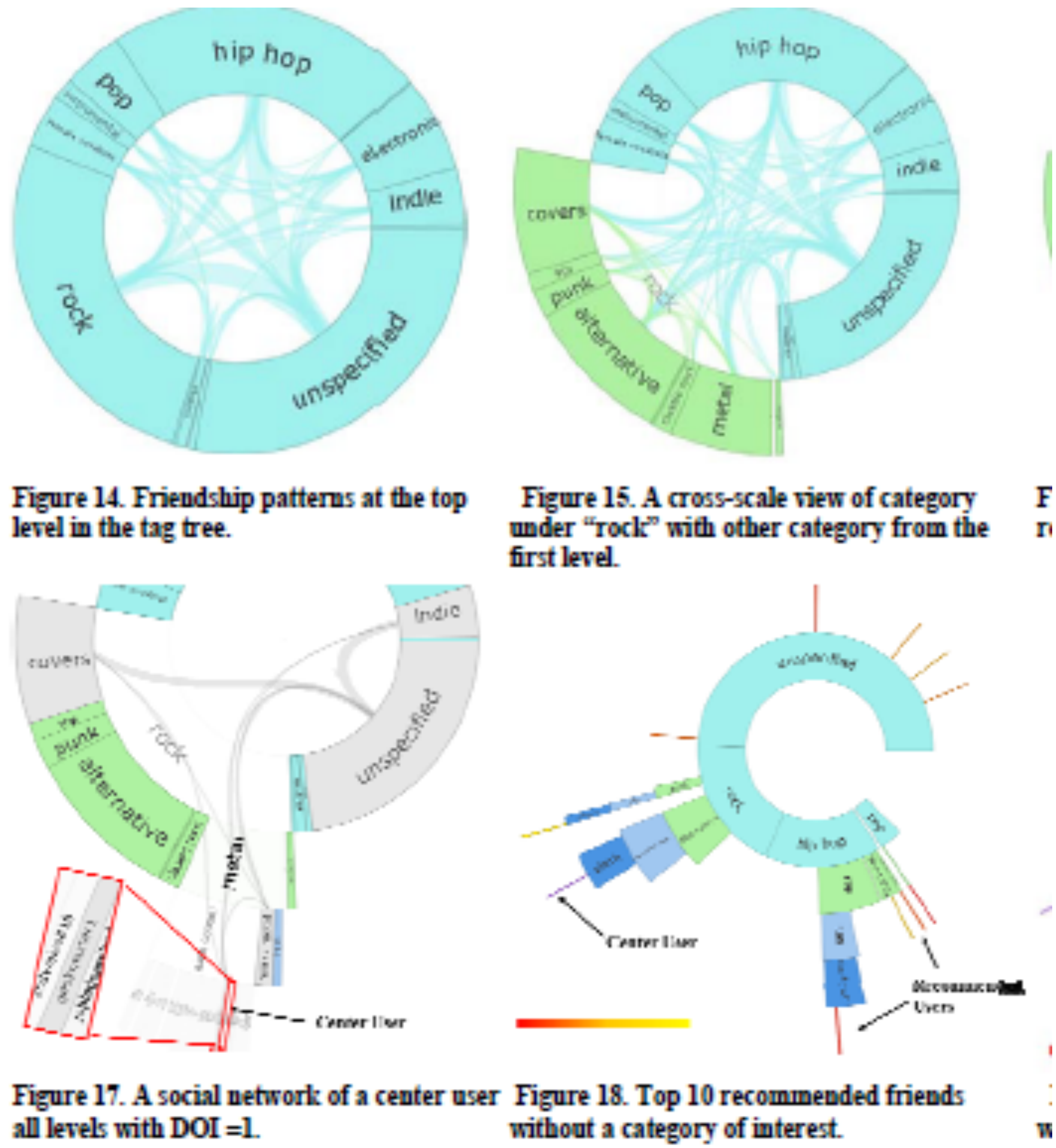


Figure 14. Friendship patterns at the top level in the tag tree.

Figure 15. A cross-scale view of category under "rock" with other category from the first level.

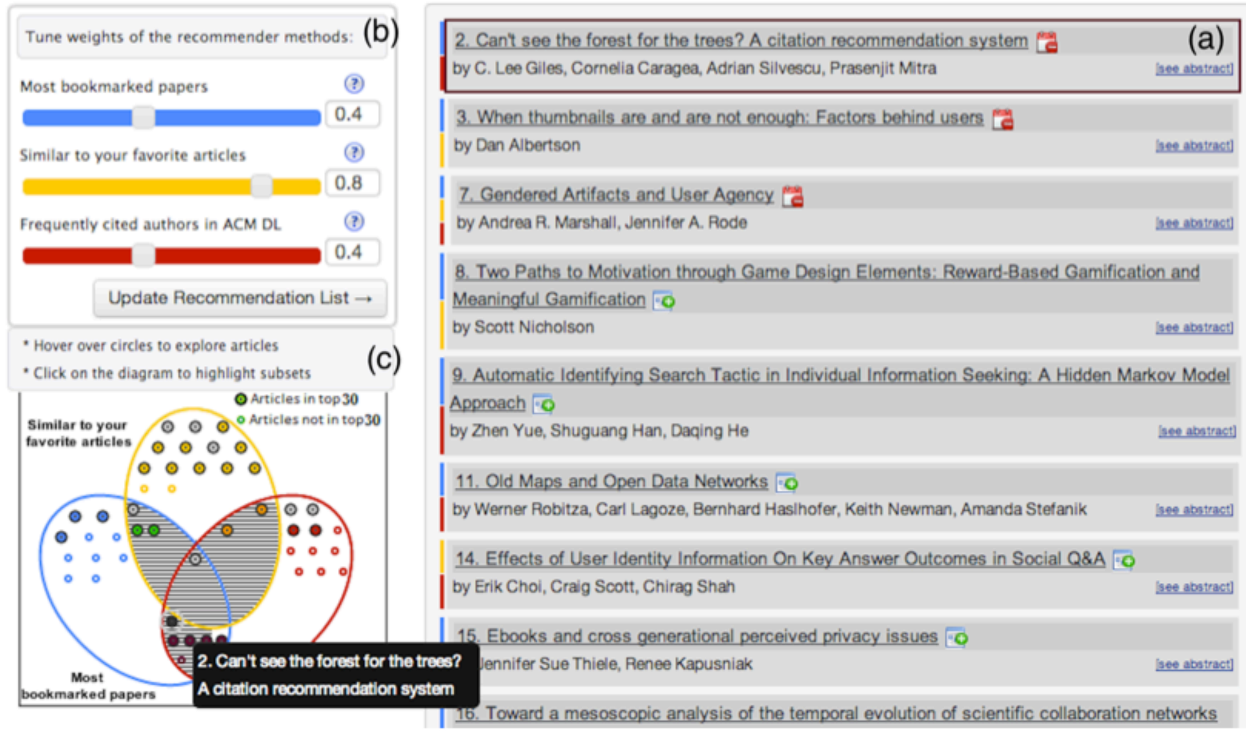
Figure 17. A social network of a center user all levels with DOI =1.

Figure 18. Top 10 recommended friends without a category of interest.

Related work on Visual RS - 6

- SetFusion

- Denis Parra, Peter Brusilovsky, and Christoph Trattner. 2014. See what you want to see: visual user-driven approach for hybrid recommendation (IUI 2014)



SetFusion: A Controllable Hybrid Recommender

Parra, D., Brusilovsky, P., Trattner, C.

IUI 2014, Haifa, Israel

https://www.youtube.com/watch?v=9LwSx1V6Yxk

Related work on Visual RS - 7

- Moodplay

- Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: Interactive Mood-based Music Discovery and Recommendation. (UMAP 2016)

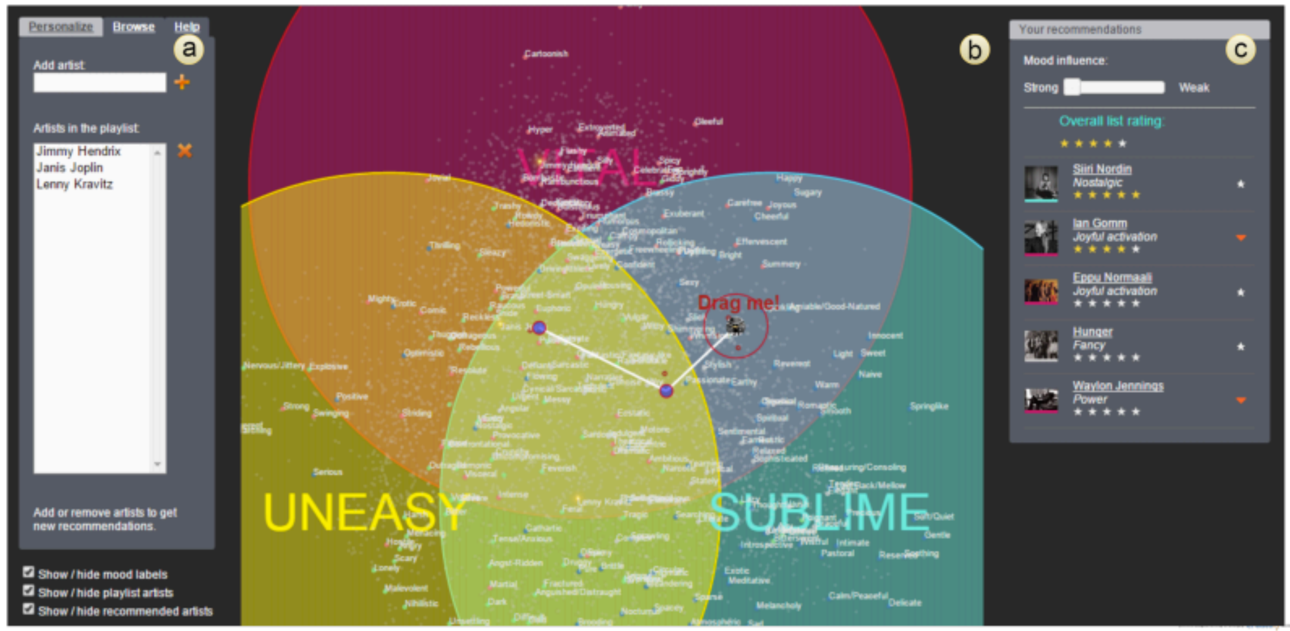


Figure 1: Screenshot of the MoodPlay interface, divided into three sections: (a) pane for entering artist names, (b) latent mood space visualization, (c) recommendation list, along with slider for adjusting mood influence

https://www.youtube.com/watch?v=eEdo32oOmcE

Controlabilidad

¿Por qué controlabilidad?

- Beyond prediction accuracy, transparency and explainability in **#recsys** have proved to be related to user satisfaction.
- Studies show an effect of controllability on user satisfaction (papers I, II, III) ~ now the details are still not completely clear
- What has not been studied?
  - Insights from our TalkExplorer studies (submitted to IUI)

Paper I

Bart P. Knijnenburg, Niels J.M. Reijmer, and Martijn C. Willemsen. 2011. **Each to his own: how different users call for different interaction methods in recommender systems**. In *Proceedings of the fifth ACM conference on Recommender systems* (RecSys '11).

Paper I

- Recommender for Energy-saving measures
- **Main message**: Controllability matters, but mainly for experts. For novices, a TopN recommendation without too much control led to better user satisfaction

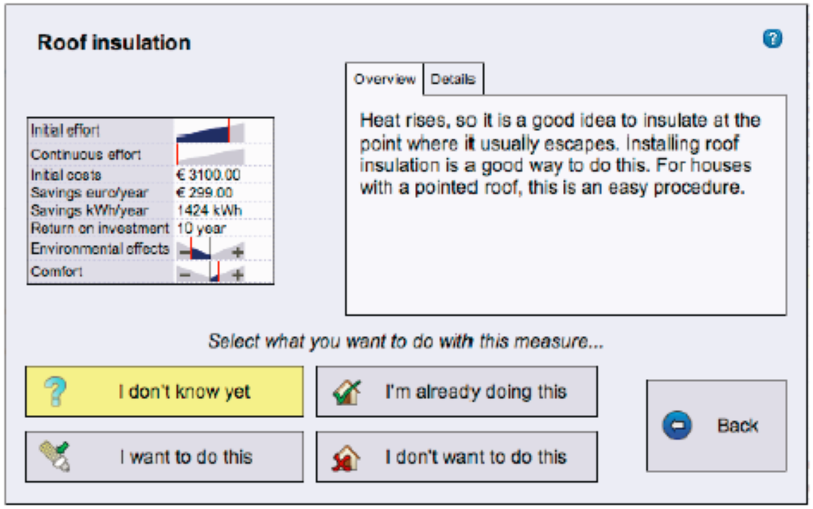


Figure 2. Screen shown to users when they click on an item

Paper II

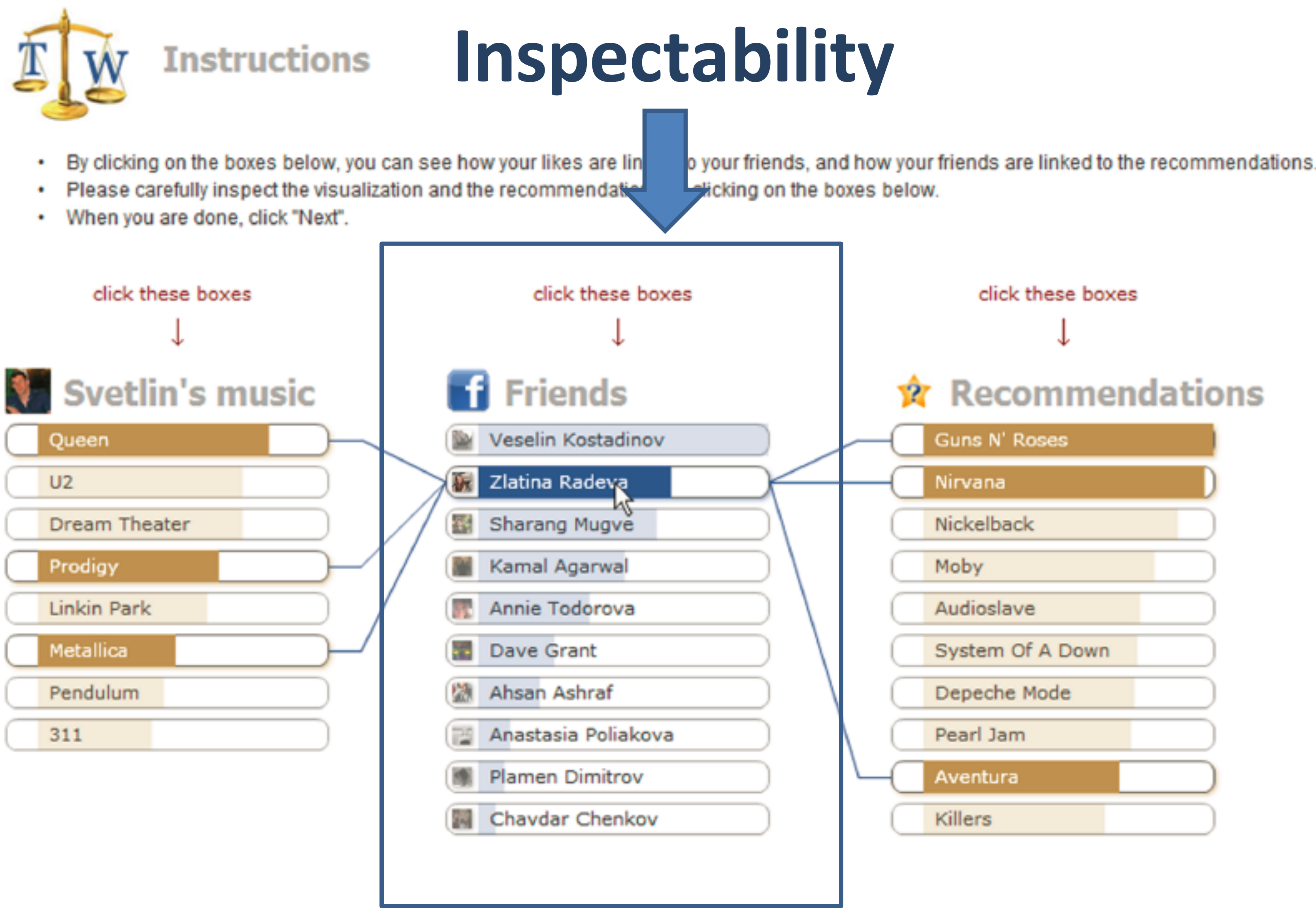
- Bart P. Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. 2012. **Inspectability and control in social recommenders**. In *Proceedings of the sixth ACM conference on Recommender systems* (RecSys '12).

Paper II

- Study on **TasteWeights**: New System introduced at RecSys 2012
- Facebook music recommender
- Gives user controls and explains how they came about
- Study with 267 (recruited in craiglist and mechanical turk)

Paper II

Inspectability



Paper II

- Summary of Results
  - Positive effects of inspectability and control, but several nuances
  - In the full graph condition, people “recognize” more recommendation, leading to better trust but lower system satisfaction (diff than recomm. Quality)
- Personal Characteristics:
  - Trusting propensity positively correlated with user satisfaction
  - Music experts feel less in control (bands to filter might be too rough) but have an overall positive perception of the system

Paper III

- Yoshinori Hijikata, Yuki Kai, and Shogo Nishida. 2012. **The relation between user intervention and user satisfaction for information recommendation**. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing* (SAC '12)

Paper III

- Terms: User Intervention instead of Control
- Study on Music Recommendation, 84 users
- Methods of user intervention
  - Rating: usual explicit feedback
  - (CI) Context Input: When / Where / With Whom
  - (CAS) Context attribute selection: country, gender, sex, unit, year
  - (PE) Profile Editing: not clear, but the highest level of intervention

Paper III

- Condition: 100 songs used for learning, 1000 for testing (experiment itself)
- 1st step: gather data from user to build recommendations
- 2nd step: randomly assign to each user 2 of the conditions: ratings, CI, CAS, PE

Paper III - results

- “”… Therefore, results show that the changes of recommendation results by user interventions improve the precision… “

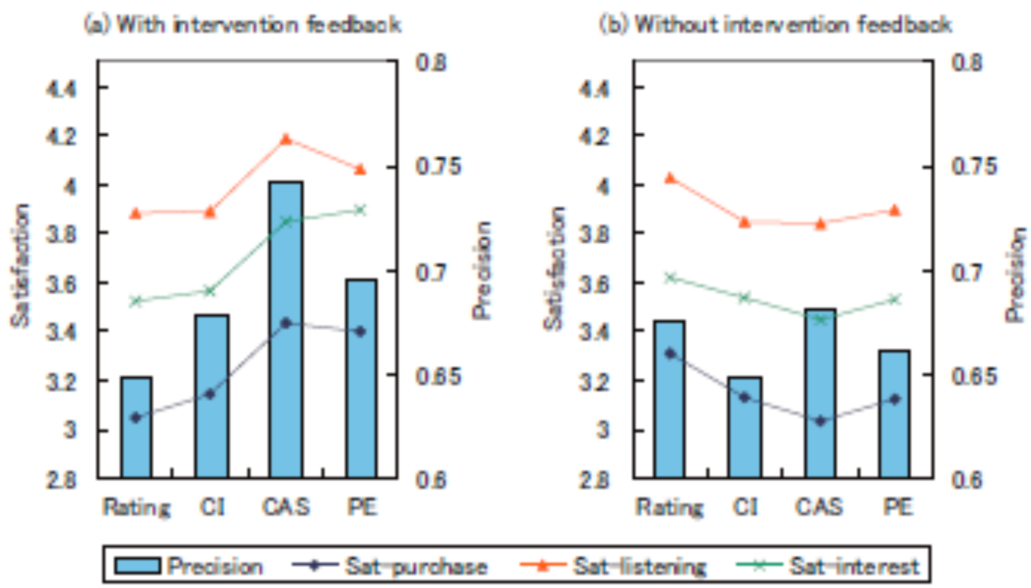


Figure 3: Relation between user intervention, precision and user satisfaction

Paper III - results

- Considering group of people with feedback effect of interest degree

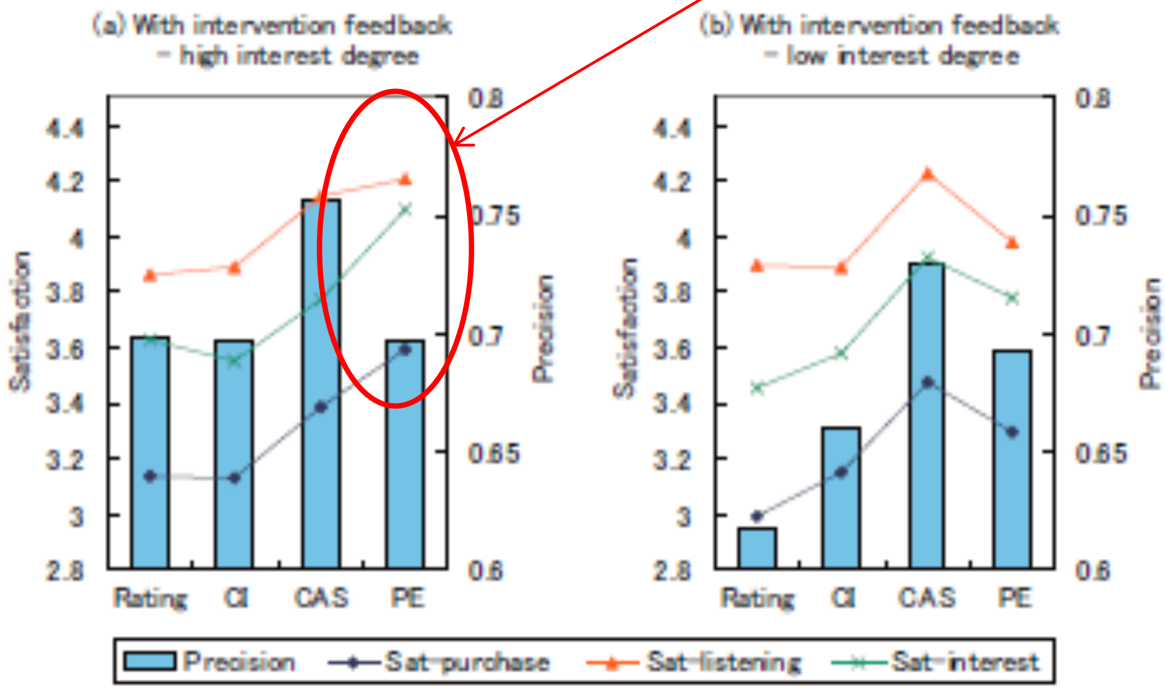


Figure 4: Relation between user intervention, precision and user satisfaction in the group with intervention feedback

Summary paper III

- When system recommends items with high precision to users with high interest in music, the more the user intervenes -> the better the user satisfaction
- NEVERTHELESS, It is still unclear whether user intervention itself influences user satisfaction

PAWS insigths

- Ahn, Jae-wook and Brusilovsky, Peter and Grady, Jonathan and He, Daqing and Syn, Sue Yeon. 2007. **Open user profiles for adaptive news systems: help or harm?** WWW 2007
- Verbert, Parra, Brusilovsky. 2013. **Visualizing Recommendations to Support Exploration, Transparency and Controllability**

Talk Explorer

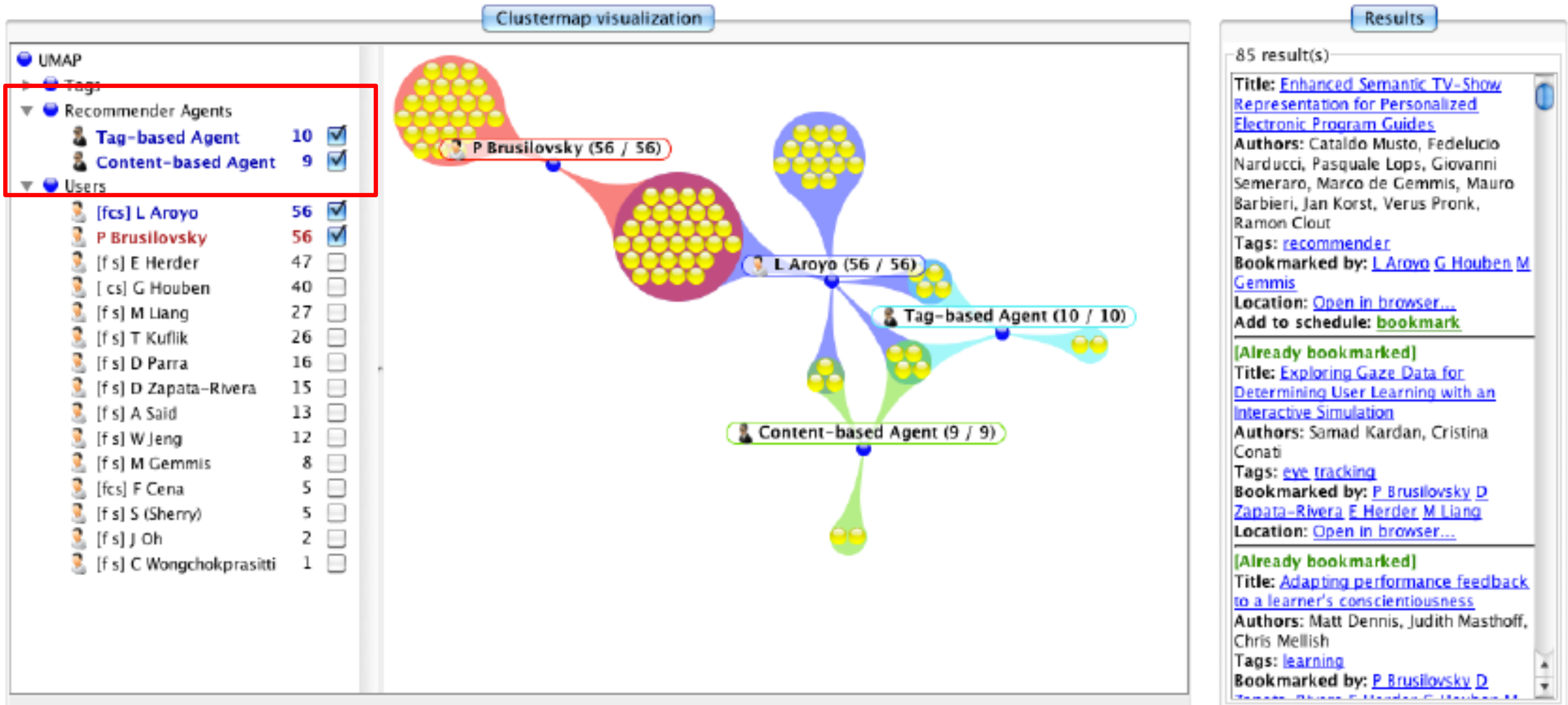
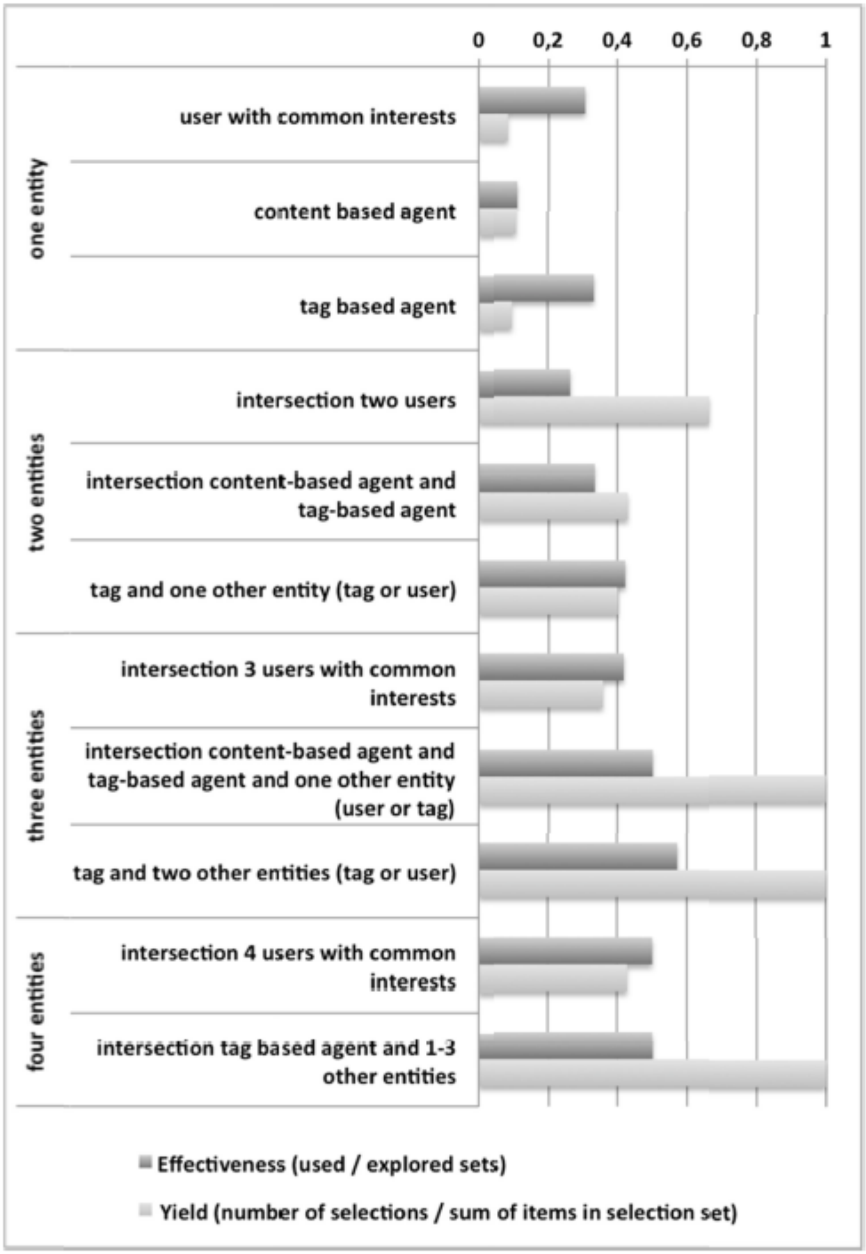
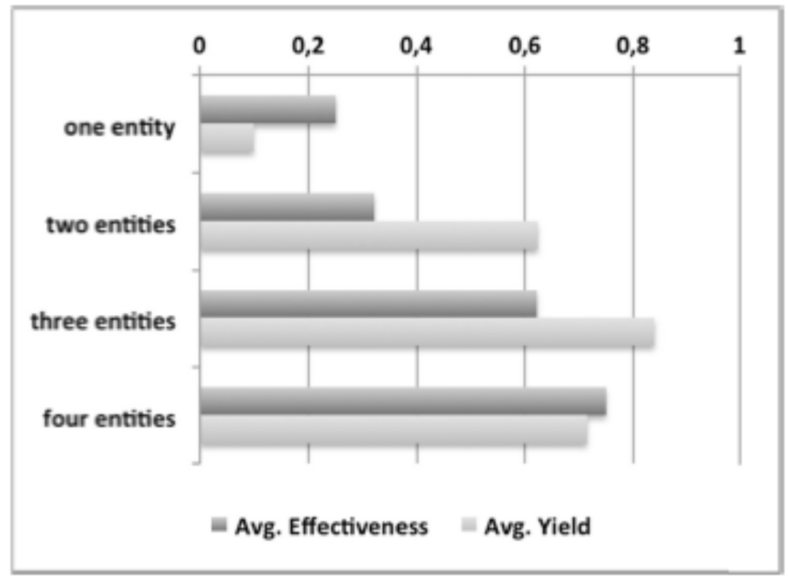


Figure 2: TalkExplorer

Talk Explorer



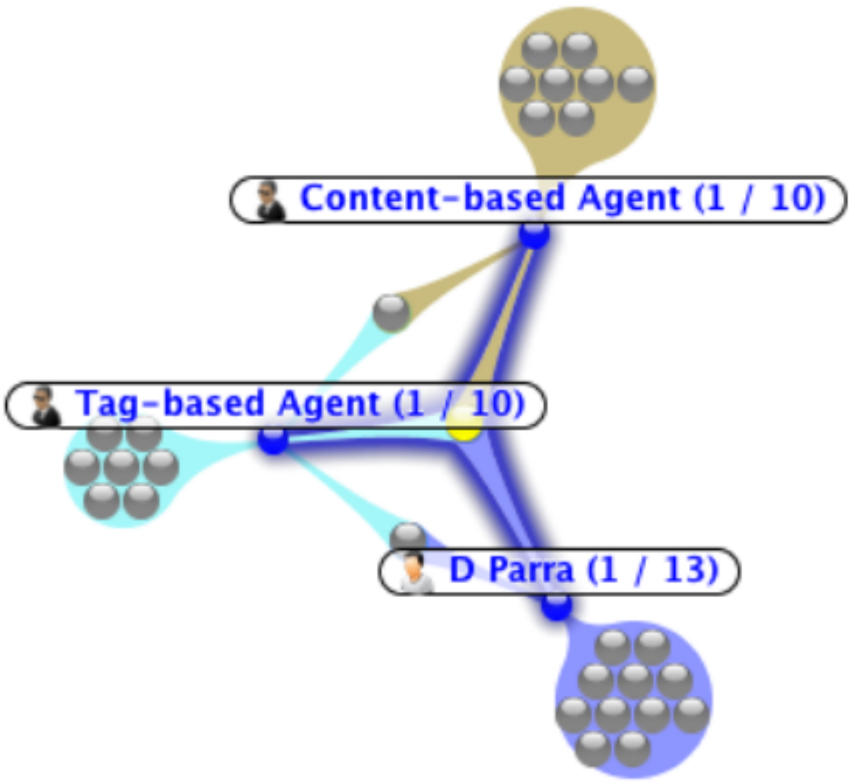


(b)

SetFusion vs. TalkExplorer

Drawback: Visualizing Intersections

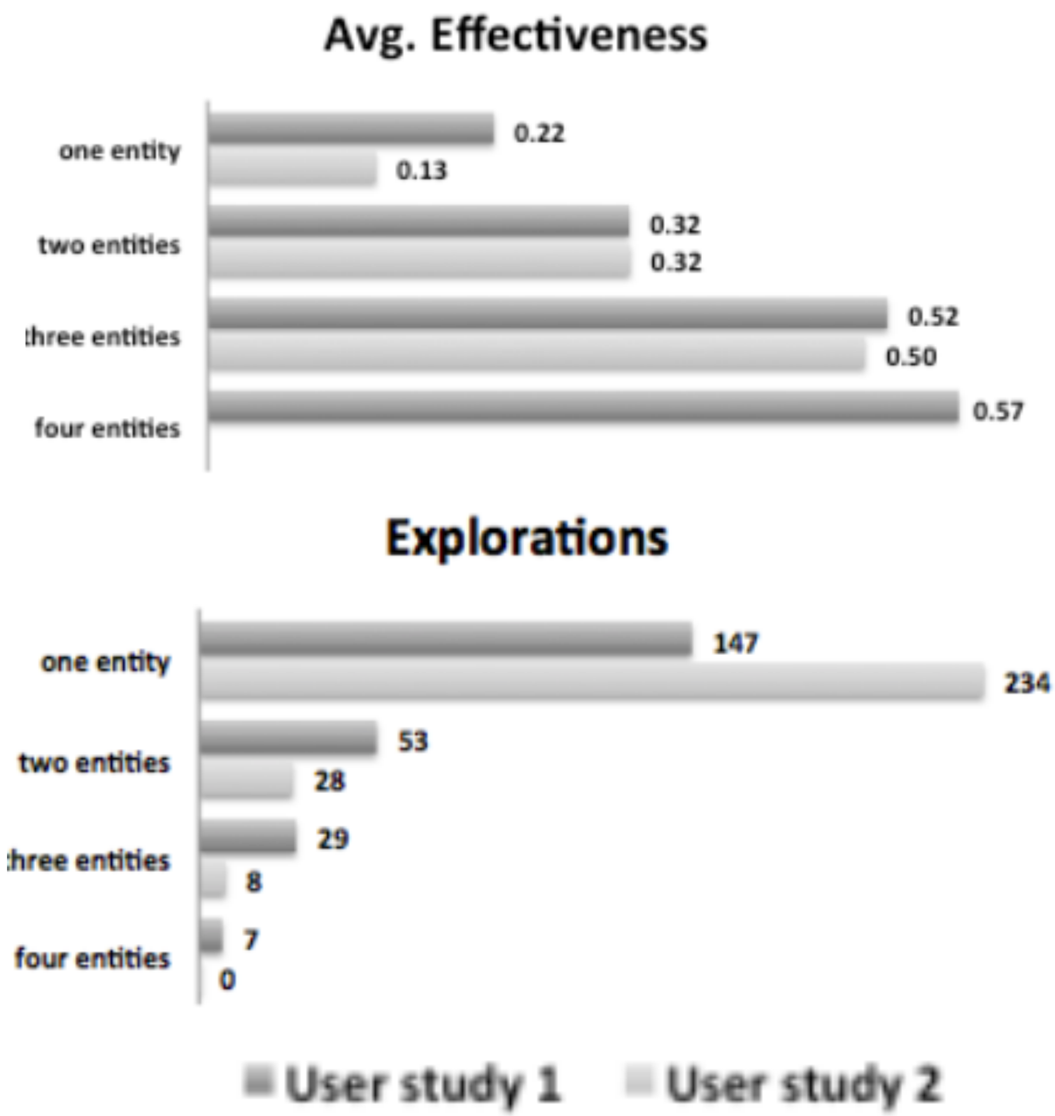
- **Venn diagram**: more natural way to visualize intersections



Clustermap

Venn diagram

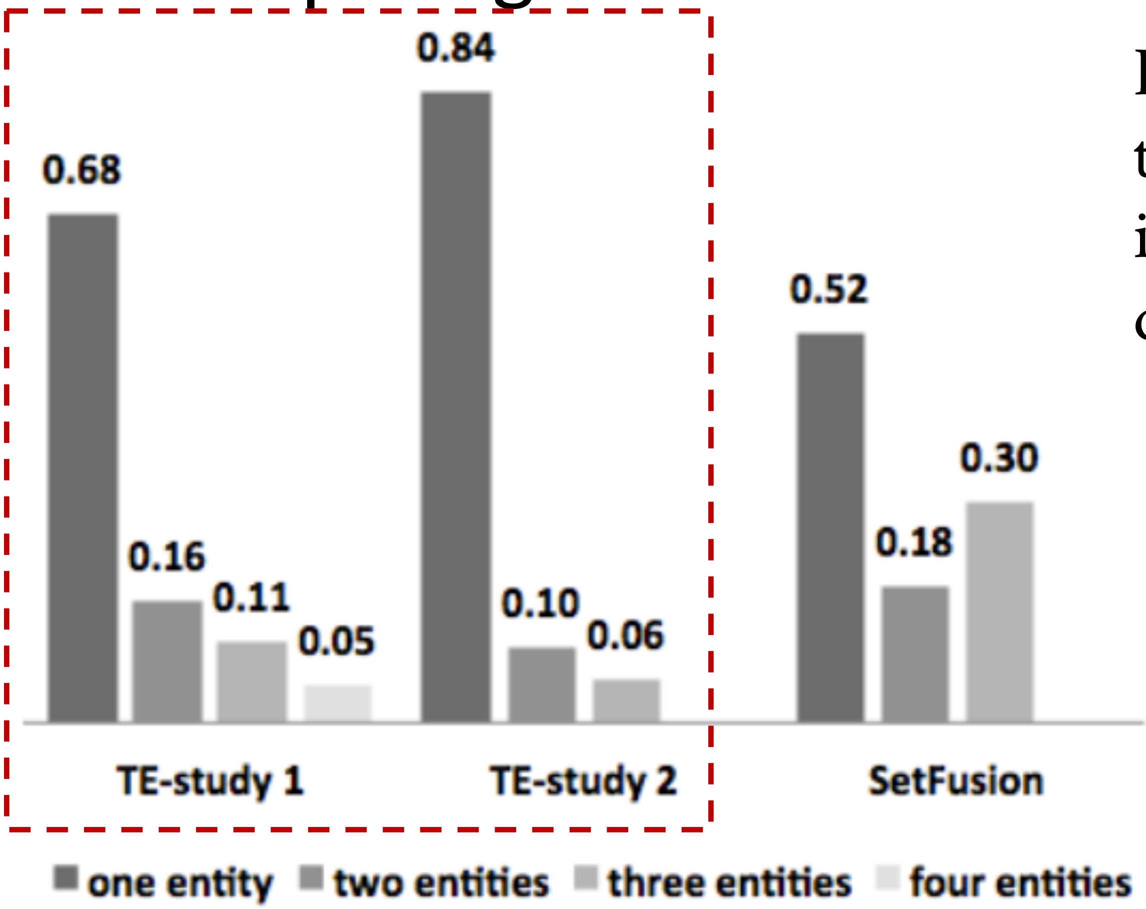
Results of Studies I & II



- Effectiveness increases with intersections of more entities
- Effectiveness wasn't affected in the field study (study 2)
- … but exploration distribution was affected

TalkExplorer vs. SetFusion

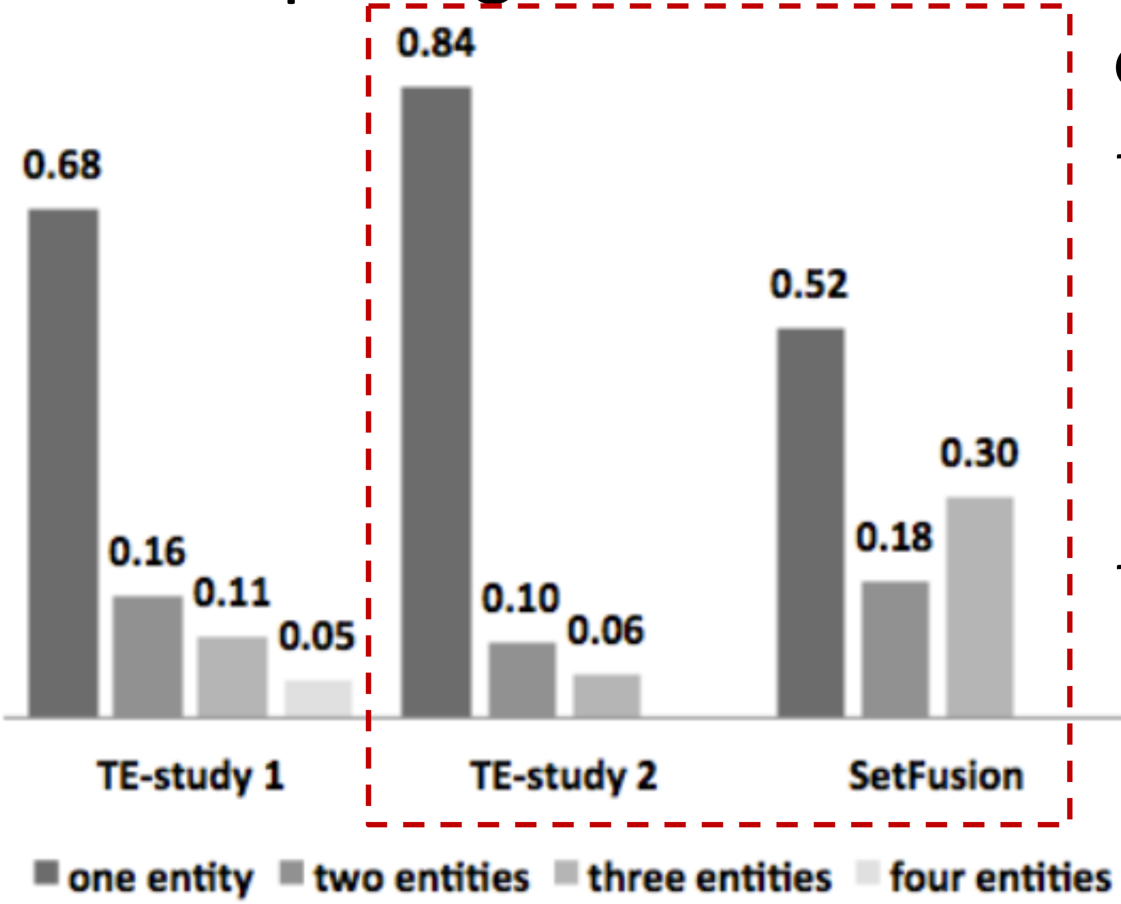
- Comparing distributions of explorations



In studies 1 and 2 over talkEplorer we observed an important change in the distribution of explorations.

TalkExplorer vs. SetFusion

- Comparing distributions of explorations



Comparing the field studies:

- In TalkExplorer, 84% of the explorations over intersections were performed over clusters of 1 item
- In SetFusion, was only 52%, compared to 48% (18% + 30%) of multiple intersections, diff. not statistically significant

Cheers!

@denisparra