



DETC2014-34767

Discovering Next Generation Product Innovations By Identifying Lead User Preferences Expressed Through Large Scale Social Media Data

Tuesday, August 19th , 2014

Suppawong Tuarob

Conrad S. Tucker

Computer Science and
Engineering

Engineering Design and Industrial and
Manufacturing Engineering

The Pennsylvania State University
{suppawong,ctucker4}@psu.edu



PRESENTATION OVERVIEW



- Background
- Motivation
- Methodology
 - Extracting Product Features
 - Identifying Latent Features
 - Identifying Product Specific and Global Lead Users
- Case Study
- Results
- Conclusions
- Future Work



NEED FOR LEAD USERS

- A **lead user** is a consumer of a product who faces needs unknown to the public

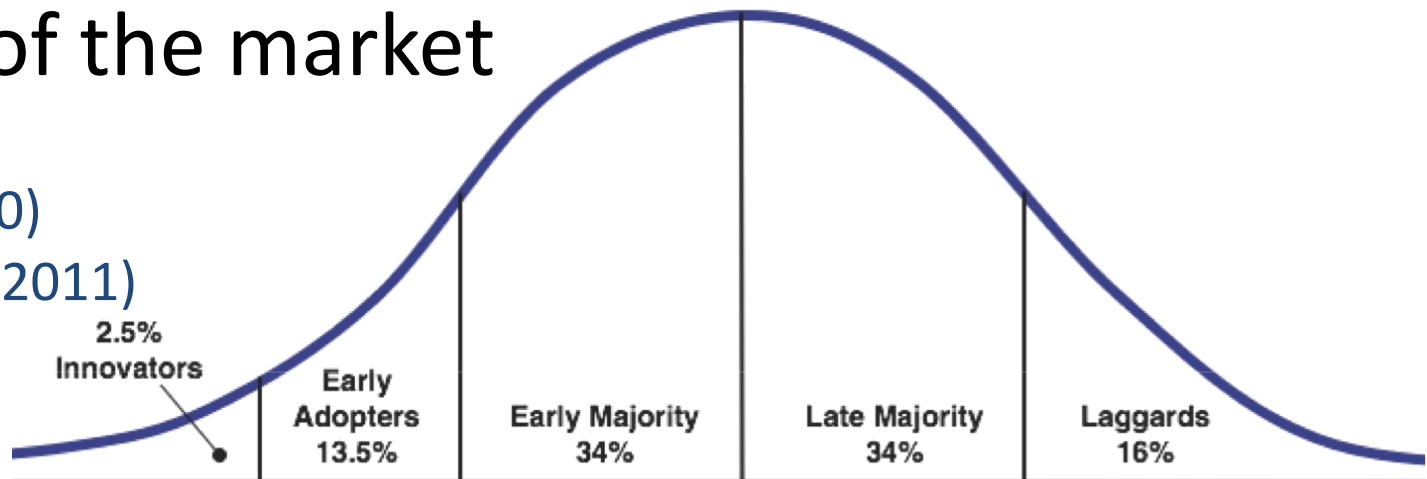
Franke et al. (2006), Schreier et al. (2007)

Lead users have two primary characteristics:

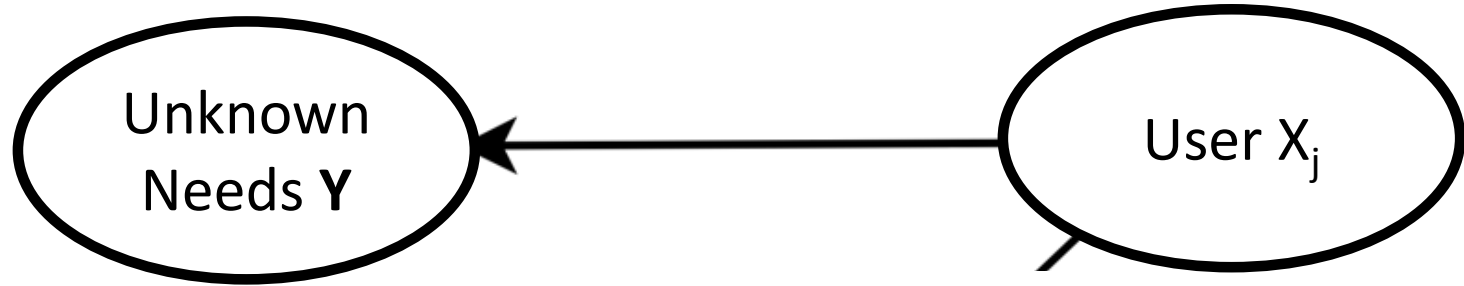
- High incentives to solve problems → Innovators
- Ahead of the market

Everett (2010)

von Hippel (2011)



ACQUIRING LEAD USER INSIGHTS



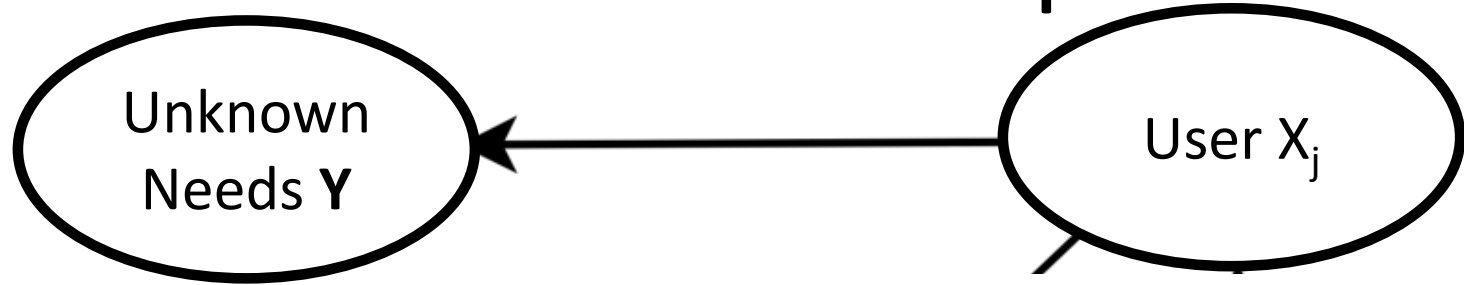
Existing Approaches: Acquire lead user knowledge through surveys, focus group interviews, or frequent interaction with the customer

-Everett (2010), von Hippel (2011), Lars, et al (2011)

CHALLENGES OF EXISTING TECHNIQUES

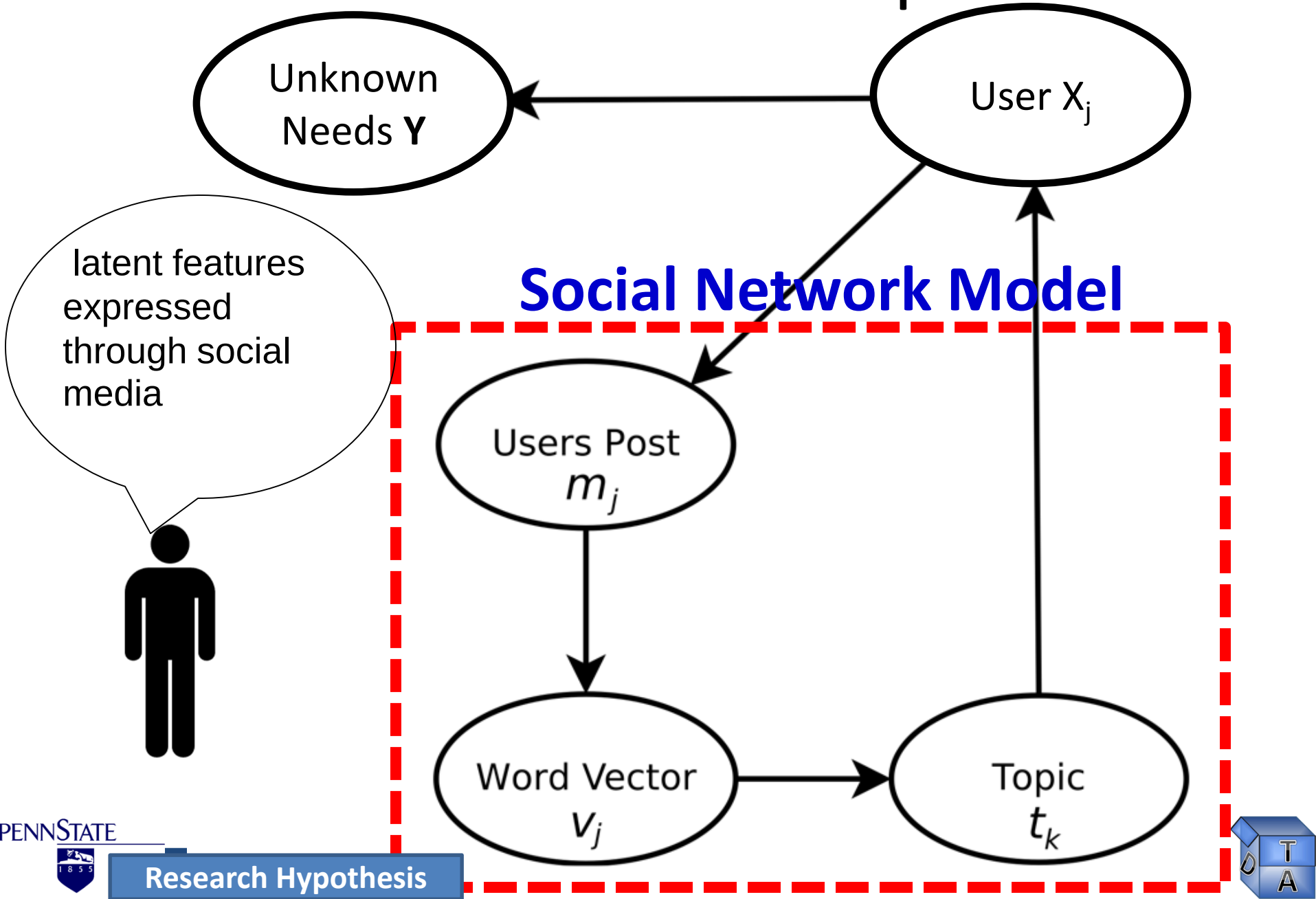
- Cost and scalability challenges of acquiring lead user insights
- Limited incentive to participate in a lead user study (from the user's perspective)

Lead User Need Acquisition

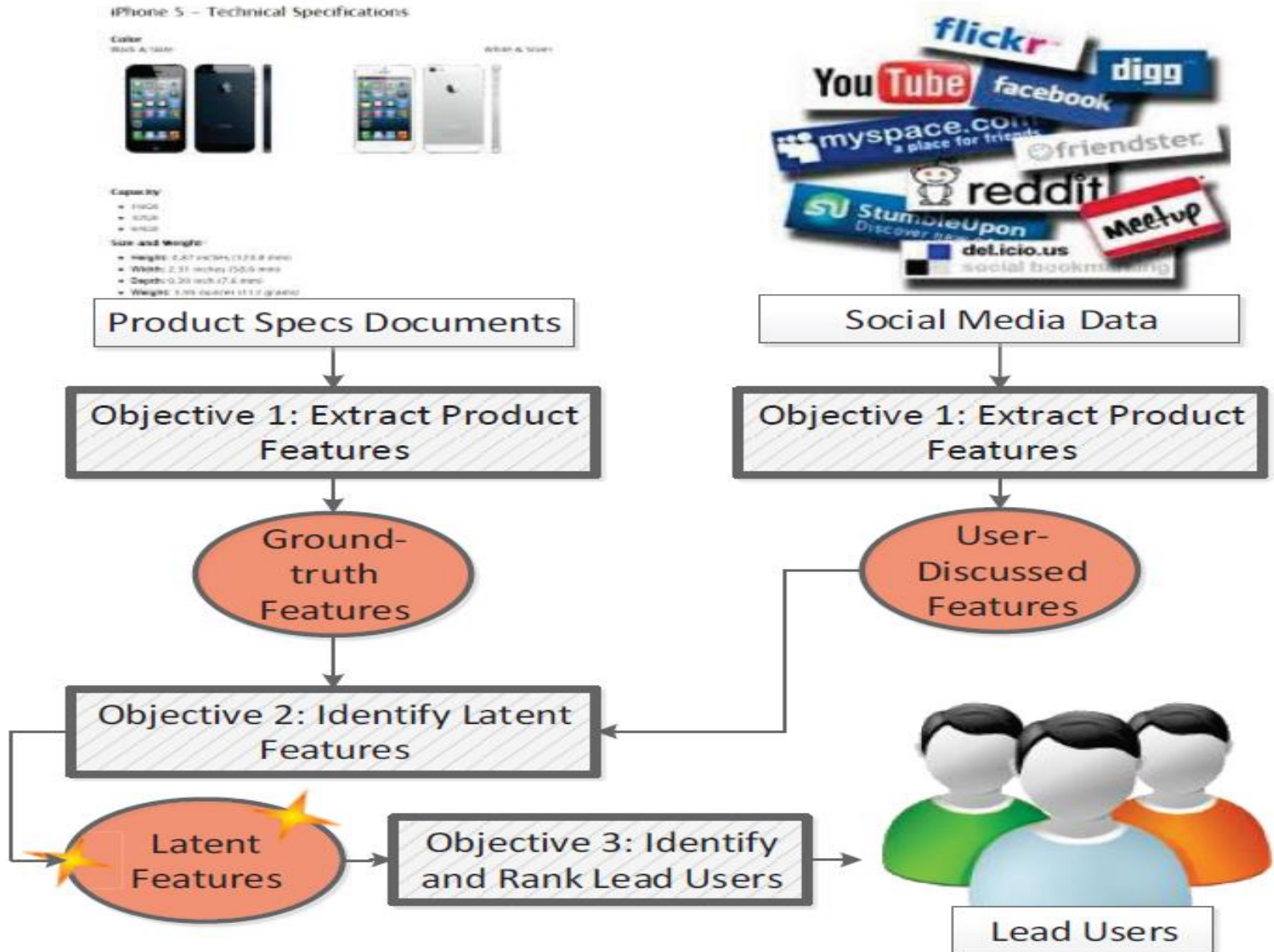


Hypothesis: topics expressed through social media networks approximate lead user needs with high accuracy

Lead User Need Acquisition



METHODOLOGY



Extracting Product Features

Algorithm 1: The feature extraction algorithm from a collection of documents

Input: D : Set of free-text documents to extract product features.

Output: E : Set of extractions. Each $e \in E$ is a tuple of $\langle \text{feature}, \text{opinion}, \text{frequency} \rangle$, for example $e = \langle \text{'onscreen keyboard'}, \text{'fantastic'}, 5 \rangle$

```
1 preprocessing;
2 for  $d \in D$  do
3     Clean  $d$  ;
4     POS tag  $d$  ;
5     Extract multi-word features ;
6 end
7 initialization;
8  $E = \emptyset$  ;
9  $T = \emptyset$  ;
10  $F = \text{Seed Features}$  ;
11 while  $E$  can still grow do
12     Learn templates from seed features;
13     Add new template to  $T$ ;
14     foreach  $d \in D$  do
15         foreach Sentence  $s \in d$  do
16              $e \leftarrow$  Extract potential feature-opinion pair using  $T$ ;
17             Add  $e$  to  $E$  ;
18         end
19     end
20     Update  $F$ ;
21 end
22  $E \leftarrow$  Clustering and normalizing features ;
```

Methodology: Extract Product Features (Example)



Product Specification Document

The rechargeable **battery** built with **lithium-ion polymer** with a charge capacity of **1440mAh**[8] is built in and cannot be replaced by the user; it is rated at **≤225 hours of standby time** and **≤8 hours of talk time**.

Extract

battery
lithium-ion polymer
1440mAh
225Hr of standby time
8Hr talk time



Social Media Message

U know with all the **glass** in the iPhone 4 they really should think about integrating a **solar panel** to recharge the **battery**.

Extract

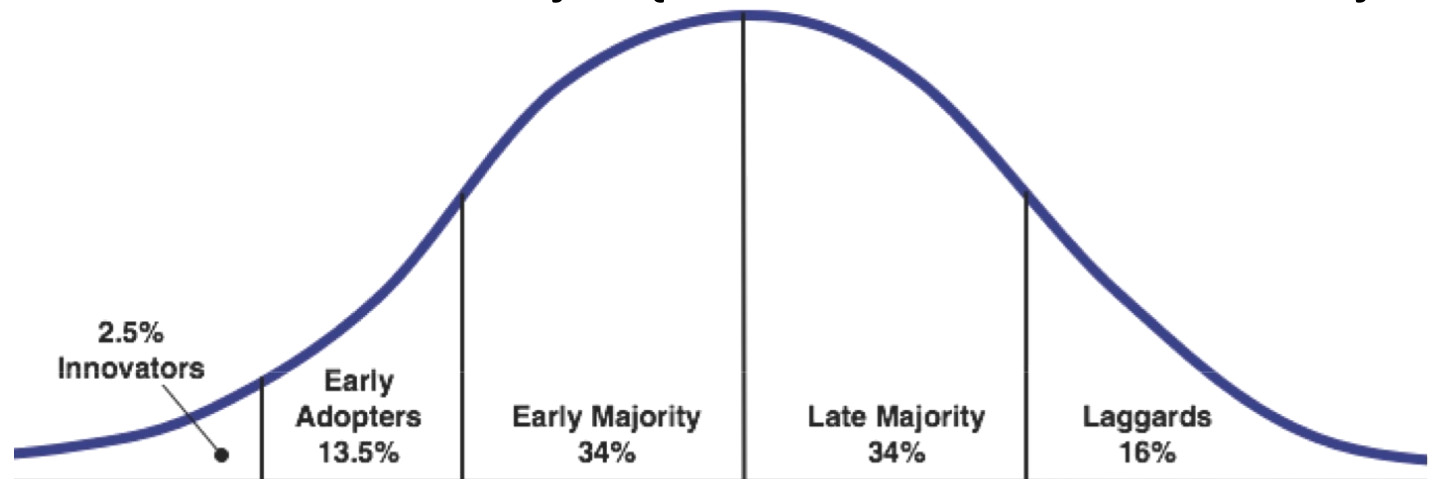
glass
solar panel
battery



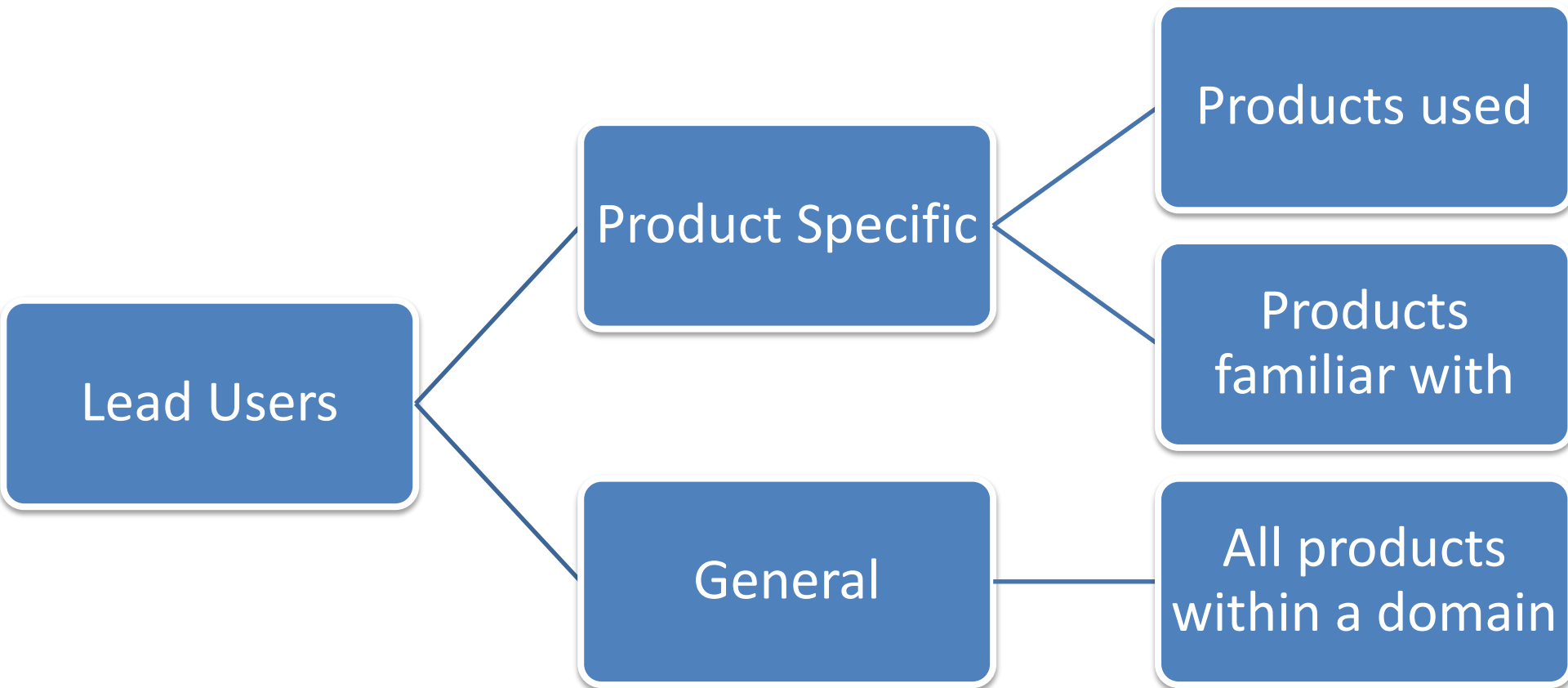
Methodology: Identify Latent Features

*“A **latent feature**: a feature that has not yet been implemented in any products within its domain”*

- Naïvely, $\{\text{Latent Features}\} = \{\text{All Features}\} - \{\text{Ground Truth Features}\}$.
- Retrieve *meaningful* latent features. So $\{\text{Latent Features}\} = \{\text{Social Discussed Features}\} - \{\text{Ground Truth Features}\}$.



Methodology: Identify Lead Users (Product Specific vs. Global)



Methodology: Identify Latent Features

- $\mathbb{S} = \{s_1, s_2, \dots, s_n\}$: a product domain
- $F^*(\mathbb{S})$ set of global latent features
- $F^*(s)$: the set of product specific latent features of product s_i

$$FF(f, F^*) = 0.5 + 0.5 \cdot \frac{|Frequency(f)|}{\sum_{f' \in F^*} |Frequency(f')|}$$

$$IPF(f, \mathbb{S}) = \log \frac{|\mathbb{S}|}{|\{s \in \mathbb{S} : f \in s\}|}$$

$$FF - IPF(f, F^*, \mathbb{S}) = FF(f, F^*) \cdot IPF(f, \mathbb{S})$$

Methodology: Identify Product Specific Lead Users

“*Product specific lead users* emit innovative ideas about the products that they use or are familiar with”

Algorithm 2: Algorithm for identifying and ranking product specific lead users of a particular product s

Input: $s \in \mathbb{S}$: The product. U : The set of all users.
 $F(G_s)$: Ground-truth features. $F(M_s)$: User discussed features. $F^*(s)$: Latent features.

Output: Ranked list of users with respect to $P(u|s)$

```
1 initialization;  
2  $I = \emptyset$  ;  
3 foreach user  $u \in U$  do  
4    $M_u \leftarrow$  The messages posted by  $u$ ;  
5   Compute  $F(M_u)$  using Algorithm 1 ;  
6   iScore  $\leftarrow$  Compute  $P(u, s)$ ;  
7   Add  $\langle u, iScore \rangle$  to  $I$ ;  
8 end  
9  $I \leftarrow$  Rank users in  $I$  by iScores;  
10 return  $I$ 
```

$$P(u|s) = \sum_{f \in F(M_u)} P(u|f, s) \cdot P(f|s)$$

$$P(u|f, s) = \begin{cases} 1 & ; f \in F^*(\mathbb{S}) \\ 0 & ; \text{Otherwise} \end{cases}$$

$$P(f|s) = \frac{1}{|F(G_s) \cup F(M_s)|}$$

Methodology: Identify Global Users

“*Global lead users* have critical, innovative ideas about all the products within the product domain.”

$$P(u) = \sum_{s \in \mathbb{S}} P(u|s) \cdot P(s)$$

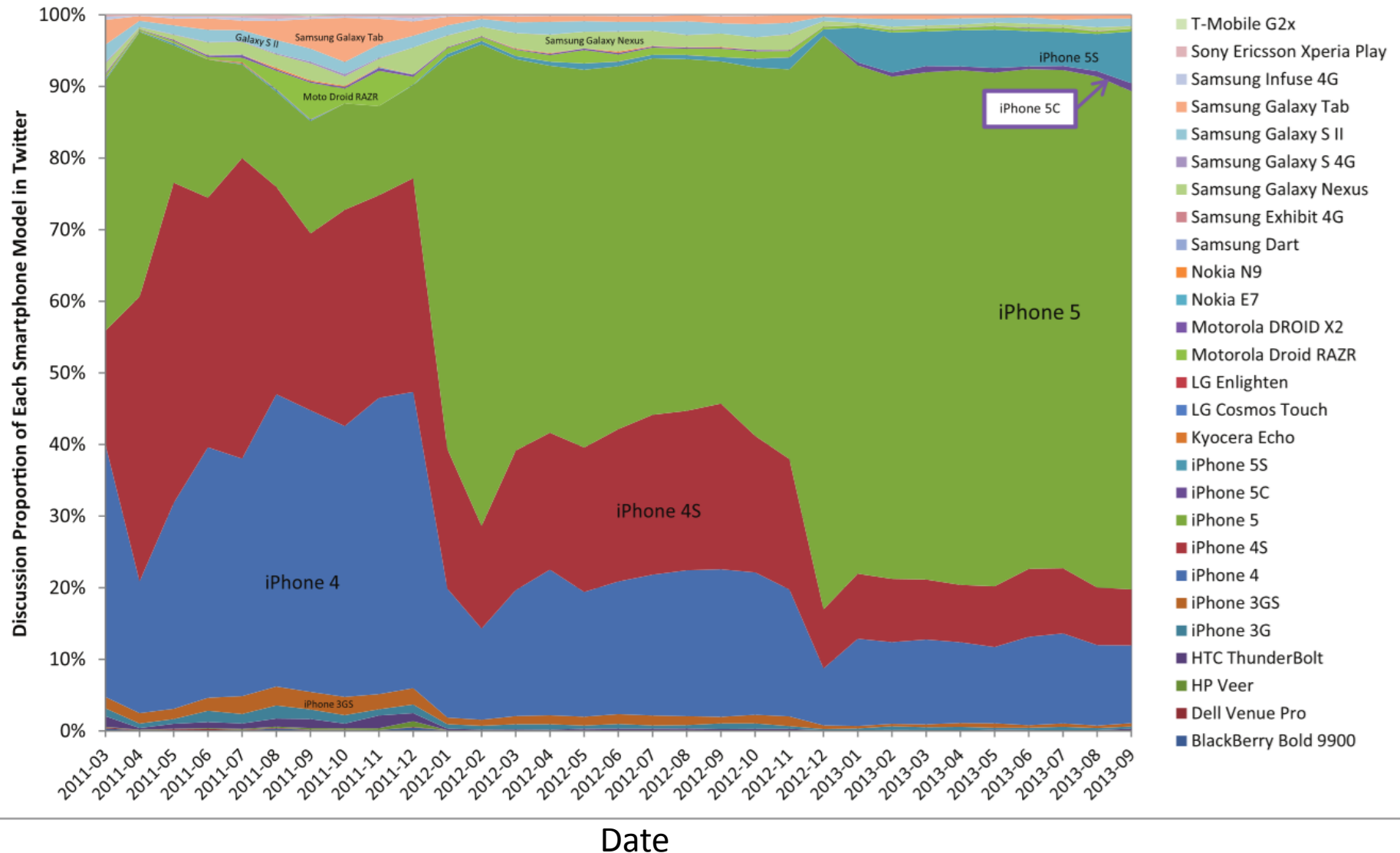
$$P(s) = \frac{|Positive(s)|}{\sum_{s' \in \mathbb{S}} |Positive(s')|}$$

Positive(s) is the set of positive messages associated with the product s.

Case Study: Smart phones and Twitter Users

- 27 smart phone products
 - Smartphone Specification Product Specification Manuals
 - Product Related Twitter Data:
 - 2.1 billions tweets in the United States during the period of 31 months, from March 2011 to September 2013.
 - Preprocess by cleaning and mapping sentiment level (positive, neutral, negative).

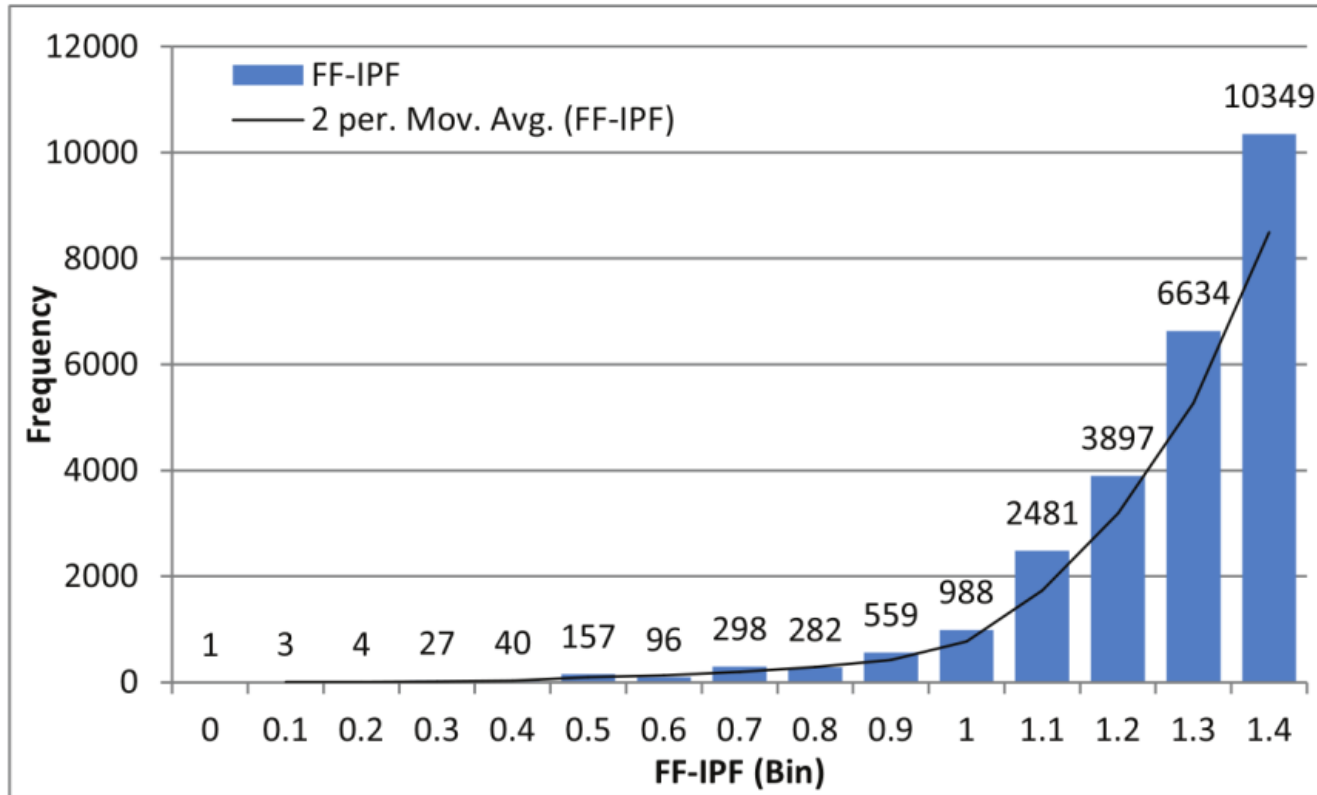
Monthly distribution of Twitter discussion of each smart phone model across the 31 month period of data collection



Results: Extract Product Features

Model	# Base Features	# User Features	# Latent Features
BlackBerry Bold 9900	1126	126	101
Dell Venue Pro	497	50	36
HP Veer	1206	76	56
HTC ThunderBolt	627	335	281
iPhone 3G	1330	532	420
iPhone 3GS	891	775	652
iPhone 4	995	6057	5720
iPhone 4S	963	5922	5582
iPhone 5	1020	13493	13050
iPhone 5C	895	833	717
iPhone 5S	973	1962	1740
Kyocera Echo	895	22	16
LG Cosmos Touch	769	11	6
LG Enlighten	1084	5	1
Motorola Droid RAZR	582	593	496
Motorola DROID X2	504	162	138
Nokia E7	749	14	10
Nokia N9	745	83	62
Samsung Dart	1178	10	6
Samsung Exhibit 4G	1331	10	7
Samsung Galaxy Nexus	456	1147	1017
Samsung Galaxy S 4G	1322	62	37
Samsung Galaxy S II	1319	801	662
Samsung Galaxy Tab	771	884	762
Samsung Infuse 4G	1121	85	60
Sony Ericsson Xperia Play	726	132	102
T-Mobile G2x	945	39	23

Results: Identify Latent Features



Histogram showing the distribution of the Feature Frequency-Inverse Product Frequency (FF-IPF) scores of 25,816 total extracted global latent features.

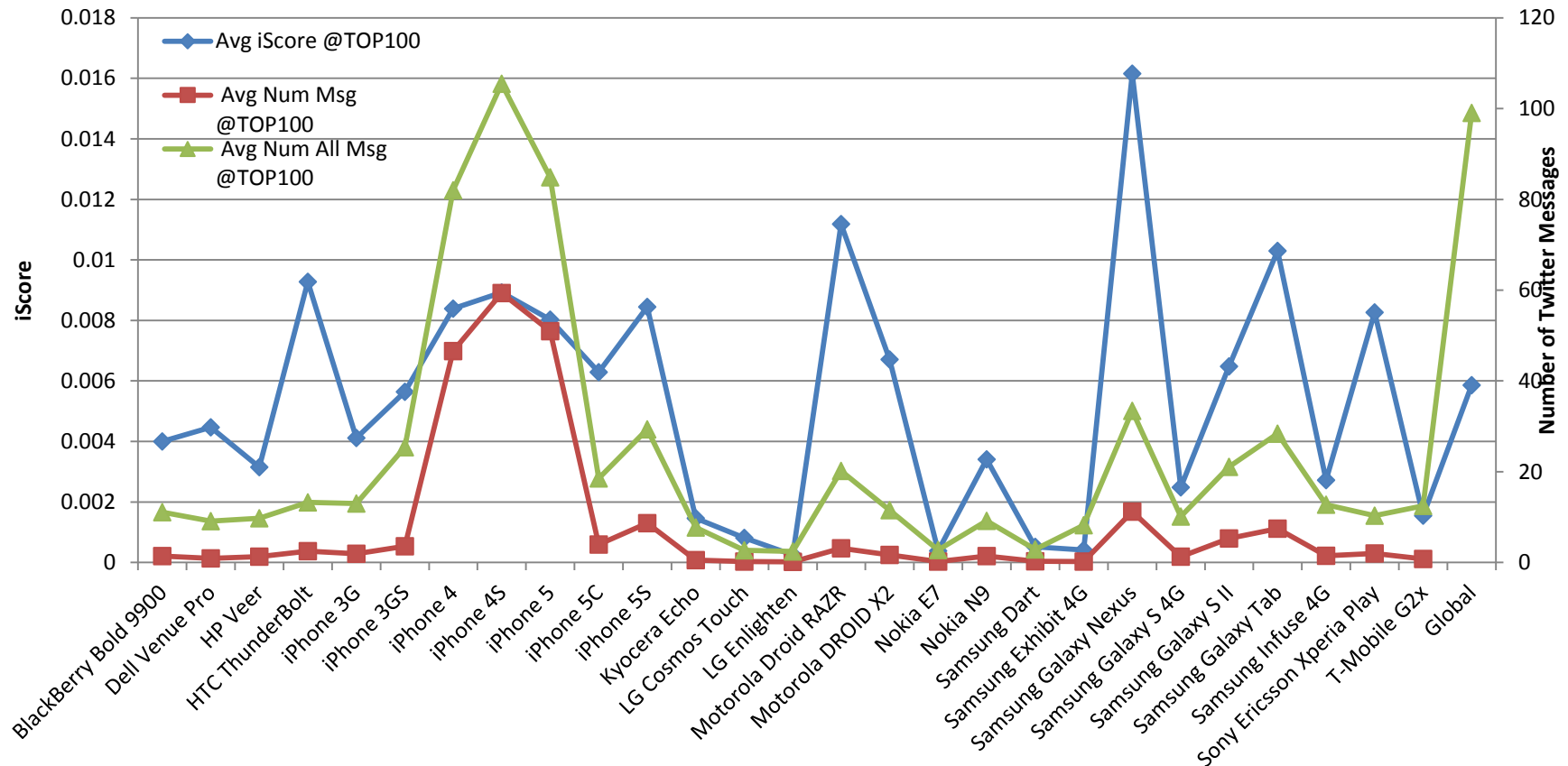
- A set of 25,816 global latent features are extracted from the smart phone related social media data.
- Latent features with FF-IPF scores lower than 1.1 are treated as noise and eliminated.

Results: Identify Latent Features

Latent Feature	FF-IPF	Example
Waterproof	1.3087	I hope Apple incorporates some of that new waterproof technology in the iPhone 5 iPhone 5 better be waterproof , shockproof, scratchproof, thisproof, thatproof, and all the rest of the proofs for \$800
Solar Panel	1.3061	... and what else would make the iPhone 5 even better, built in solar power charging ! U know with all the glass in the iPhone 4 they really should think about integrating a solar panel to recharge the battery.
Hybrid	1.3027	I wish there was an #android phone out there that was a hybrid of the best features on the droid razr maxx and the galaxy nexus . I need a hybrid-iPhone 4s so the battery can hold on all day when I'm at #vmworld. Steve, are you listening? :)
Tooth Pick	1.3023	I hope iPhone 5 borrows from Swiss Army and finally adds a removable tooth pick .
iHome	1.3021	My life would be 827492916 times better if my iHome took my iPhone 5 First world problem: mad because my iPhone 5 is not compatible with this iHome dock in the hotel room.

Top 5 latent features across the chosen smart phone models, FF-IPF scores, and example tweets that related to the latent features

Results: Identify Lead Users



Average product specific iScore (i.e. $P(u|s)$) and global iScore (i.e. $P(u)$) of top 100 lead users across all the selected smartphone models, along with average numbers of tweets both related to each smartphone model (Avg Num Msg @TOP100) and average number of tweets related to smartphone in general (Avg Num All Msg @TOP100)

Results: Identify Lead Users

Model	Product iScore	Sample Twitter Message
Samsung Galaxy Nexus	0.0496	I wish there was an #android phone out there that was a hybrid of the best features on the droid razr maxx and the galaxy nexus.
HTC ThunderBolt	0.0308	HTC Thunderbolt #fail: Connect phone to PC to access drivers on included SD card ... but need drivers installed to access SD card from PC
iPhone 5	0.0174	but unless Siri can do more than just talk ...I'm not sold! #iPhone5
Sony Ericsson Xperia Play	0.0085	Hmm.. Playing games supporting Xperia Play controls. Wish I could use PS3 controller .. Makes me want an LTE Xperia Play with Tegra3..
Kyocera Echo	0.0077	Kyocera Echo needs to develop its own apps .

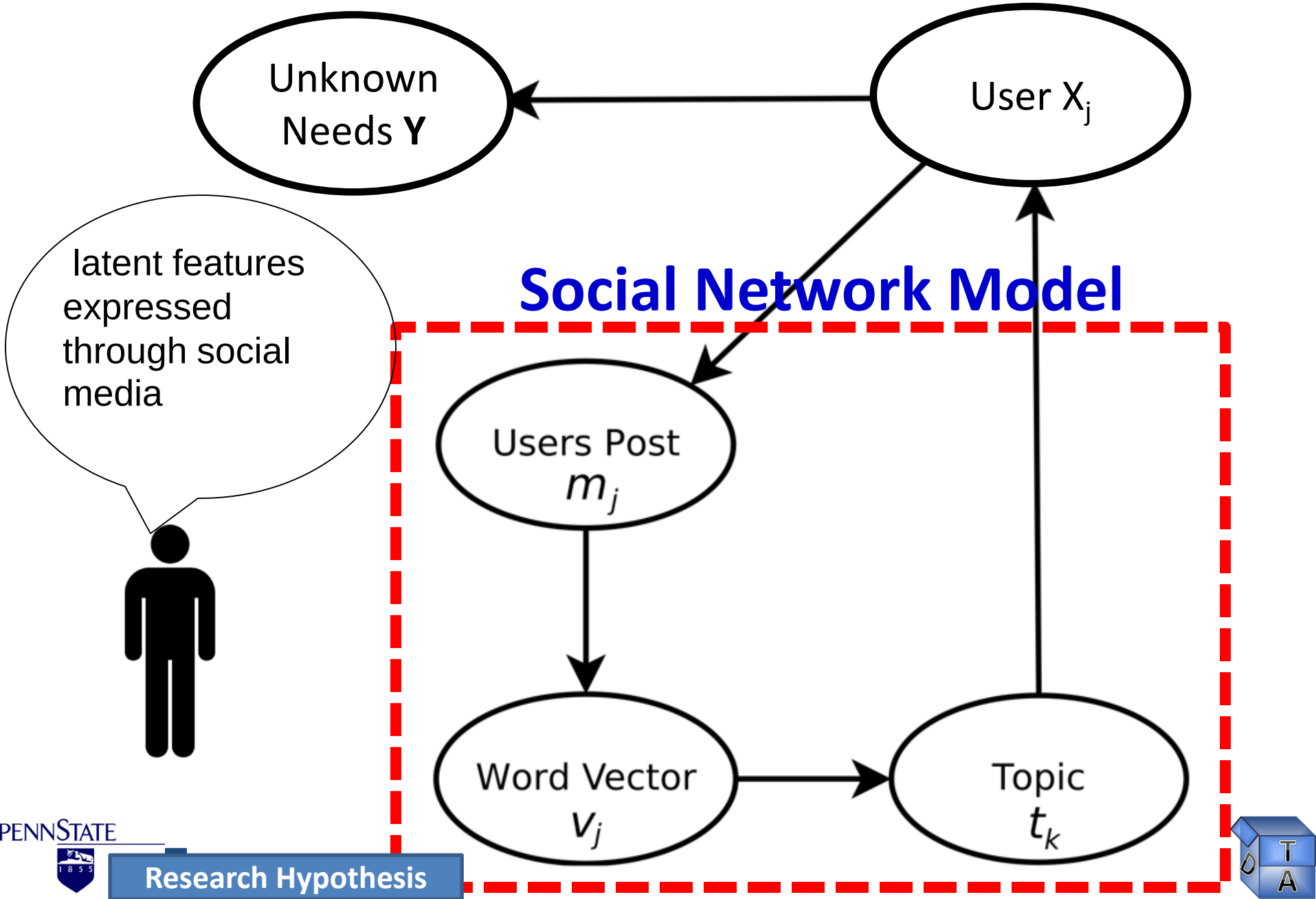
Sample tweets from the top lead user of each sample five smart phone models. These tweets suggest product innovative improvement for each corresponding product.

Results: Identify Lead Users

Global iScore	Sample Twitter Message
0.0127	I wish there were a tweak for the iPhone 4S that would <i>indicate "4G" instead of just 3G</i> when I'm connected with a HSDPA+ connection.
0.0126	If you trust my instinct, the iPhone 5S will come in <i>multiple colors and two display sizes</i>
0.0113	Very exciting Siri on the iPhone 4S <i>activates when you "raise it to your ear"</i> that'd b awesome.
0.0107	I wish i could use my iPhone as a <i>universal remote</i> control.
0.0105	Since iPhone already does fingerprint, Sumsung should <i>scan eyes</i> .

Sample tweets from the top 5 global lead users of the smart phone domain. These tweets suggest product innovation.

Conclusion and Path Forward



Acknowledgement & References

Contributors:

- D.A.T.A. Lab: Suppawong Tuarob, Conrad Tucker

References

- [1] What is big data?—Bringing big data to the enterprise. <http://www-01.ibm.com/software/ph/data/bigdata/>, 2013. [Online; accessed 16 August 2013].
- [2] Toni Ahlqvist. Social media roadmaps: exploring the futures triggered by social media. VTT, 2008.
- [3] Sinan Aral and Dylan Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.
- [4] Carliss Baldwin and Eric Von Hippel. Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. Harvard Business School Finance Working Paper, (10-038):4764–09, 2010.
- [5] D.A. Batallas and A.A. Yassine. Information leaders in product development organizational networks: Social network analysis of the design structure matrix. *Engineering Management, IEEE Transactions on*, 53(4):570–582, 2006.
- [6] Pierre R. Berthon, Leyland F. Pitt, Ian McCarthy, and Steven M. Kates. When customers get clever: Managerial approaches to dealing with creative consumers. *Business Horizons*, 50(1):39 – 47, 2007.

