

Attributed network embedding

- ❑ **Motivations & challenges**


 - What are attributed networks and why embedding
 - Formal definitions and challenges

- ❑ Mining attributed networks with shallow embedding


- ❑ Mining attributed networks with deep embedding

- ❑ Human-centric network analysis

Example of node attributes

 **Texas A&M University** @TAMU · Jun 7

A new \$1 million @ENERGY grant will help @TAMUEngineering explore the use of big data, A.I., & machine learning to bolster power grids! #tamu



Big Data Analytics Could Reduce Power Grid Outages - Texas A&M T...

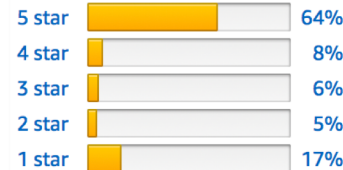
A Texas A&M team will use a \$1 million Department of Energy grant for research that could improve assessment of events that affect power sys...
today.tamu.edu

3 replies 23 likes

Customer Reviews

★★★★☆ 623

4.3 out of 5 stars



Apple 15" MacBook Pro
by Apple

Capacity: 15 Inch, 2.9GHz Intel Core i7

[Change](#)

Price: **\$2,599.00** + Free shipping

[Write a review](#)

Top positive review

[See all 450 positive reviews](#)

59 people found this helpful

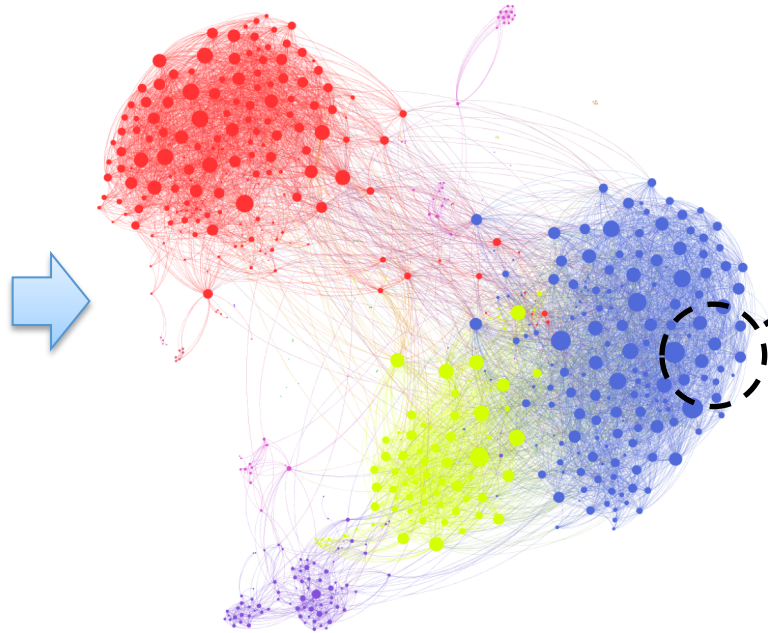
★★★★★ **It's a Macbook Pro Maxed out from 2016**

By Timothy D. Gray on January 23, 2018

Many of the negative reviews here are from people that either don't understand computers or bought during the short time the specs posted by amazon as to what people were buying were wrong. Amazon has now fixed that and what you see is now accurate.

- Examples: **user content** in social media, **reviews** in co-purchasing networks, & paper abstracts in citation networks

Attributed networks are prevalent in practice



Nodes Have Different Attributes

- Node attributes: a rich set of data describing the unique characteristics of each node

Node attributes & network are correlated

The image shows a screenshot of a Twitter profile page for Texas A&M University. At the top, the profile name is "Texas A&M University" with the handle "@TAMU" and a verified account icon. Below the name, statistics are displayed: Tweets (18.7K), Following (1,733), Followers (258K), Likes (12.3K), and Lists (8). The "Following" count is highlighted with a red underline. Below the statistics, there are three profile cards. The first card is for "Texas A&M School of Innovation" with handle "@TAMUischool". The second card is for "Academic Success Center" with handle "@SuccessTAMU". Both cards have a "Follow" button. The background of the profile cards features the Texas A&M logo and various text elements related to the university's academic success initiatives.

- Node attributes and network influence each other and are inherently correlated
 - Explained by Homophily & social influence
 - High correlation of user posts & following relationships
 - Strong association between customer reviews & co-purchasing networks

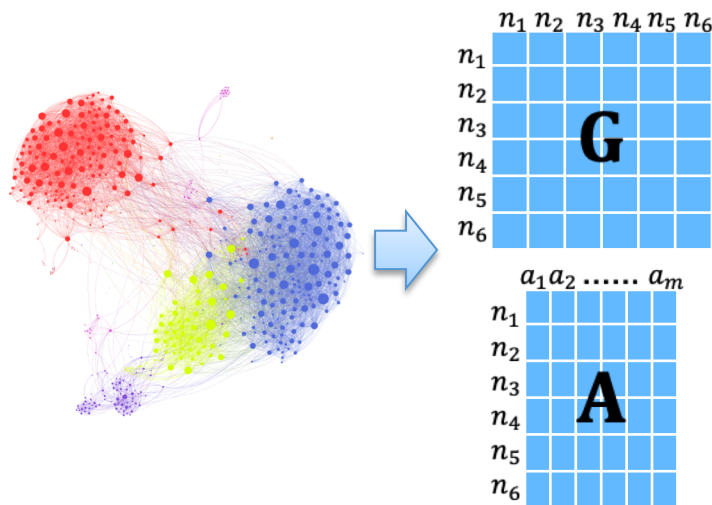
Hypothesis testing on correlation

Dataset	Scenarios	CorrCoef	p-value
BlogCatalog	Real-world	3.69e-002	0.00e-016
	RandomMean	3.14e-005	0.18
	RandomMax	1.40e-003	4.42e-016
Flickr	Real-world	1.85e-002	0.00e-016
	RandomMean	2.15e-005	0.49
	RandomMax	5.48e-004	3.37e-003

- Hypothesis: there is no correlation between network affinities and node attribute affinities (a significance level of 0.05)
- CorrCoef: Pearson correlation coefficient of two types of affinities
- Real-world network vs randomly-generated networks
 - Mean and max results of 100 synthetic networks

Attributed network embedding

Network & Node Attributes



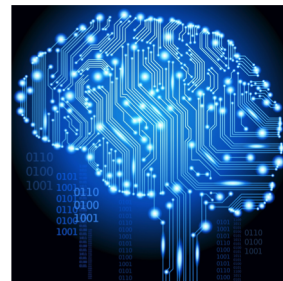
Embedding Representation

A matrix **H** representing the embedding of nodes. It is a 6x2 matrix with columns of numerical values and rows labeled n_1 through n_6 . Below the matrix is the label "← Latent Space →".

0.54	0.27	n_1
0.22	0.91	n_2
0.55	0.28	n_3
0.98	0.11	n_4
0.32	0.87	n_5
0.26	0.11	n_6

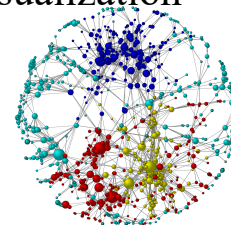
H

Off-the-shelf ML Algorithms



Tasks

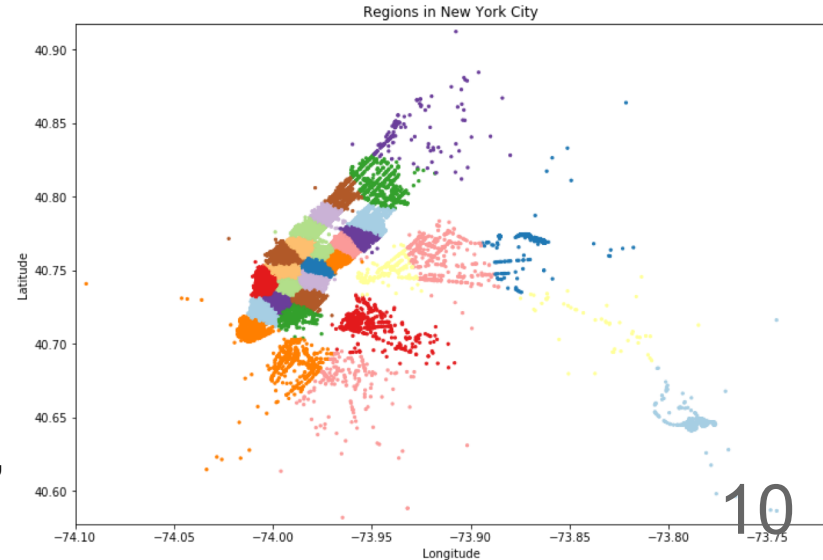
- Clustering
- Link Prediction
- Classification
- Visualization
- ...



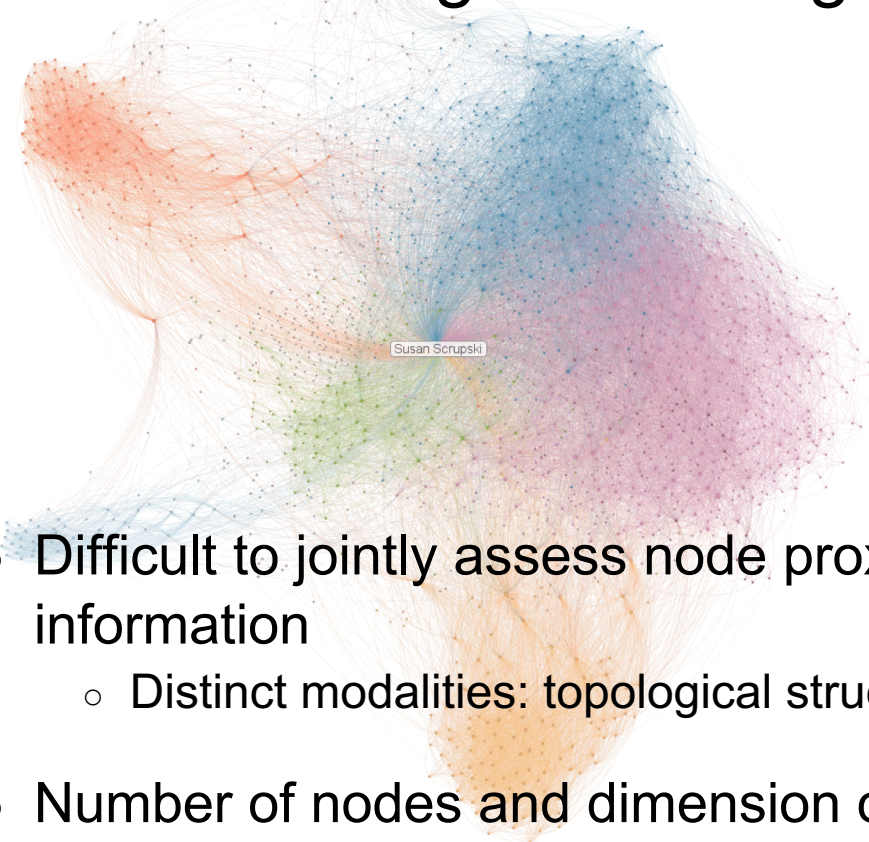
- Given **G** and **A**, we aim to represent each node as a d -dimensional vector \mathbf{h}_i , such that **H** can preserve node proximity both in network and node attributes

Why attributed network embedding

- Traditional graph theory based analysis achieves suboptimal in **large-scale networks with complex tasks**
 - Shortest path, maximum flow, centrality
- Aim to take advantage of **off-the-shelf** machine learning algorithms
- Provide **general ways** to handle the heterogeneous info in networked systems
 - Friend recommendation: social links, textual posts, categorical attributes, images.
 - Taxi demand forecast: region networks, demographic and meteorological data.



Challenges: heterogeneity & large scale



Texas A&M University @TAMU · Jun 7

A new \$1 million @ENERGY grant will help @TAMUEngineering explore the use of big data, A.I., & machine learning to bolster power grids! #tam



Big Data Analytics Could Reduce Power Grid Outages - Texas A&M T...

A Texas A&M team will use a \$1 million Department of Energy grant for research that could improve assessment of events that affect power sys...
today.tamu.edu



3

23



Texas A&M University @TAMU · Jun 7

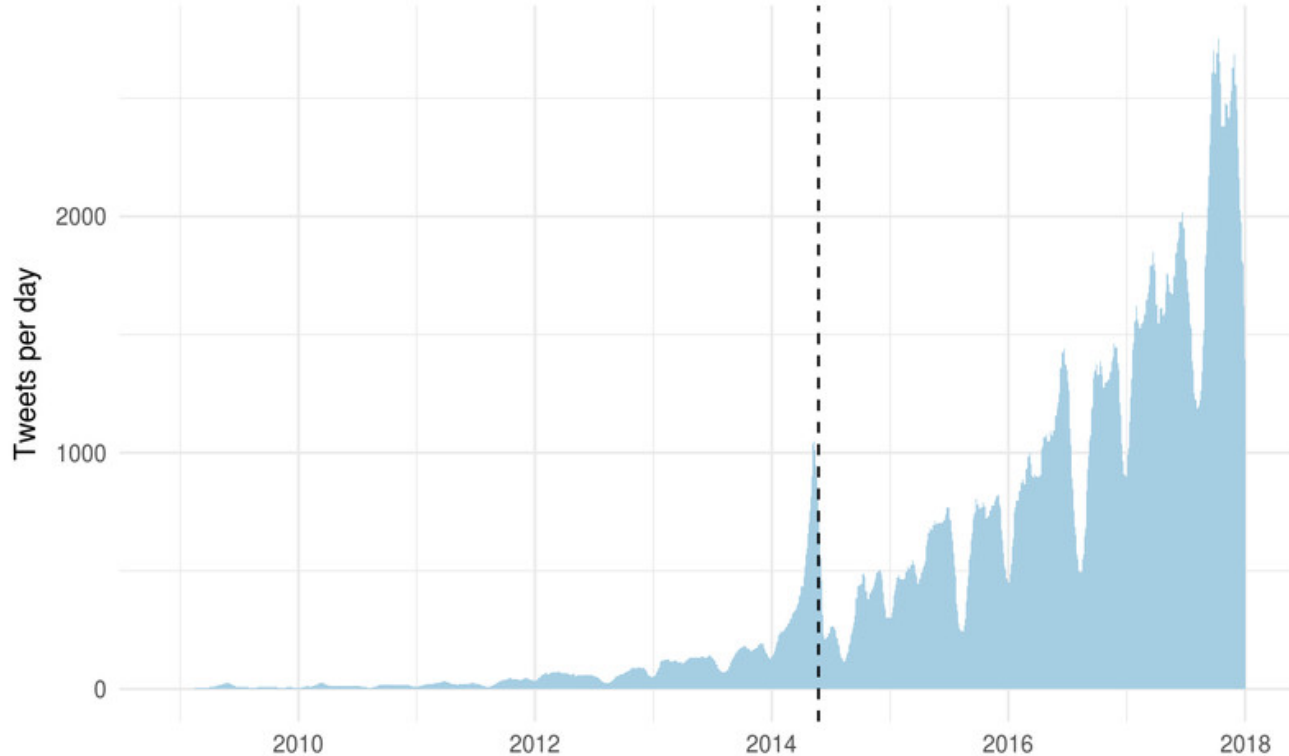
Texas A&M is ranked No. 8 in the nation in this year's @schoolsEDU 'Best Colleges' survey! Whoop! 🙌 #tam

- Difficult to jointly assess node proximity from the heterogeneous information
 - Distinct modalities: topological structures & node attributes
- Number of nodes and dimension of attributes could be large
 - It could be expensive to store or manipulate the high-dimensional matrices such as node attribute similarity

Real-world attributes are high-dimensional

Number of tweets posted by all current MEP per day. (MEP: European Parliament)

The dotted line presents the final day of the latest European Parliament elections



Data characteristics vary significantly

Product information

Capacity: 15 Inch, 2.9GHz Intel Core i7, 16GB RAM, 512GB SSD | Style: 15" w/ Touch Bar | Color: Space Gray

Technical Details [Collapse all](#)

Summary

Screen Size	15 inches
Max Screen Resolution	2880x1800 pixels
Processor	2.9 GHz Intel Core i7
RAM	16 GB DDR3 SDRAM
Hard Drive	512 GB Flash Memory Solid State
Graphics Coprocessor	Radeon Pro 560
Chipset Brand	intel
Card Description	Dedicated
Number of USB 3.0 Ports	2
Average Battery Life (in hours)	10 hours

Customer Reviews

Apple 15" MacBook Pro by Apple

Capacity: 15 Inch, 2.9GHz Intel Core i7
[Change](#)
Price: \$2,599.00 + Free shipping

Write a review

★★★★☆ 623
4.3 out of 5 stars

5 star	64%
4 star	8%
3 star	6%
2 star	5%
1 star	17%

Top positive review
[See all 450 positive reviews](#)

59 people found this helpful
★★★★★ **It's a Macbook Pro Maxed out from 2016**
By Timothy D. Gray on January 23, 2018

Many of the negative reviews here are from people that either don't understand computers or bought during the short time the specs posted by amazon as to what people were buying were wrong. Amazon has now fixed that and what you see is now accurate.

- Different types of useful heterogeneous info, such as multiple networks, multiple types of node attributes, & labels
 - Facebook: attributes in introduction, words in posts, content in photos, predefined groups etc.
 - Amazon: product info, customer reviews, customer purchasing records, customer viewing history, etc.