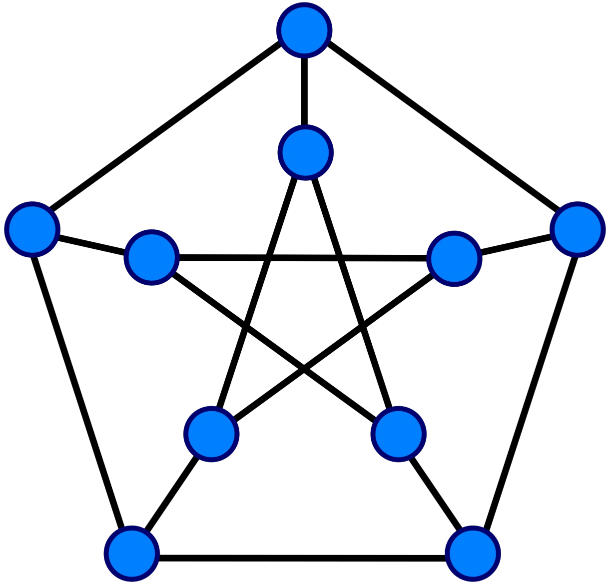
Bayesian approach in applications

Tomasz Kajdanowicz

Embedding in graphs

Graph encoding



[0, 1.2, 0.8, 4, ...]

[1, 0.1, 1.2, 3, ...]

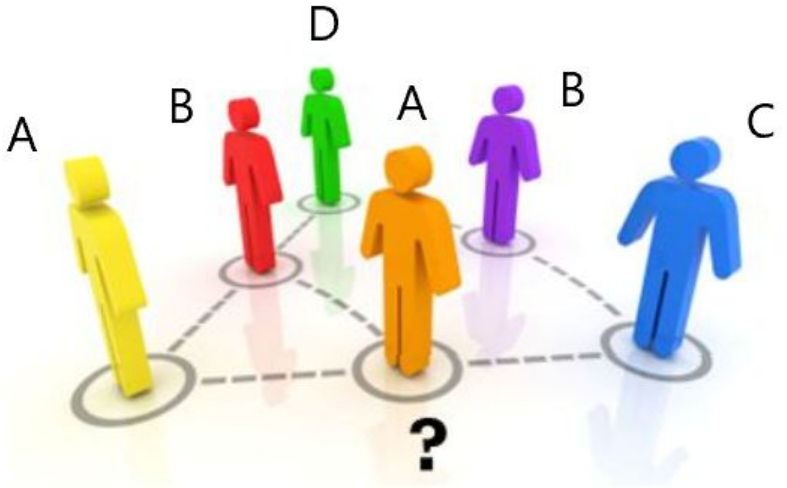
[0, 1.2, 0.5, 2, ...]

[1, 0.3, 2.2, 4, ...]

Node embedding

Applications

- node classification
- user behaviour prediction
- advertisement personalization
- friends recommendation



Methods for graph encoding

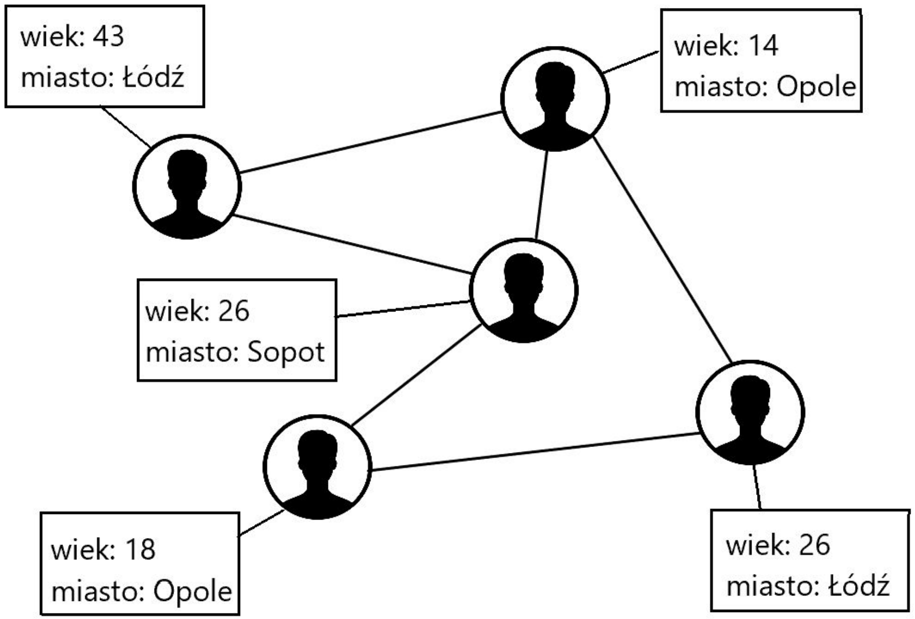
PREPROCESSING

“hard-coded” features

- features included in the dataset
- node labels

network centrality measures:

- node degree
- clustering coefficient
- number of shortest paths passing through a given node



Methods for graph encoding

REPRESENTATION LEARNING - EMBEDDING

Learning of a mapping from nodes/edges/subgraphs/graphs to a low-dimensional vector space.

d << |V|

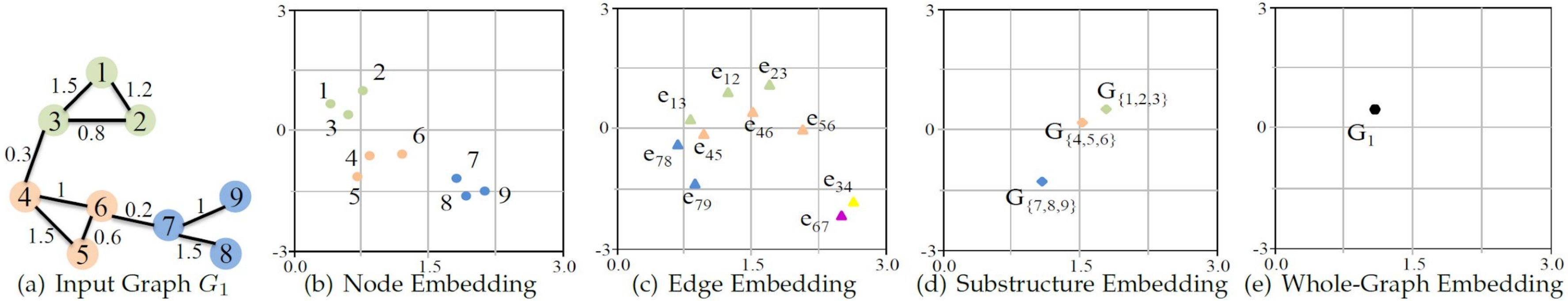
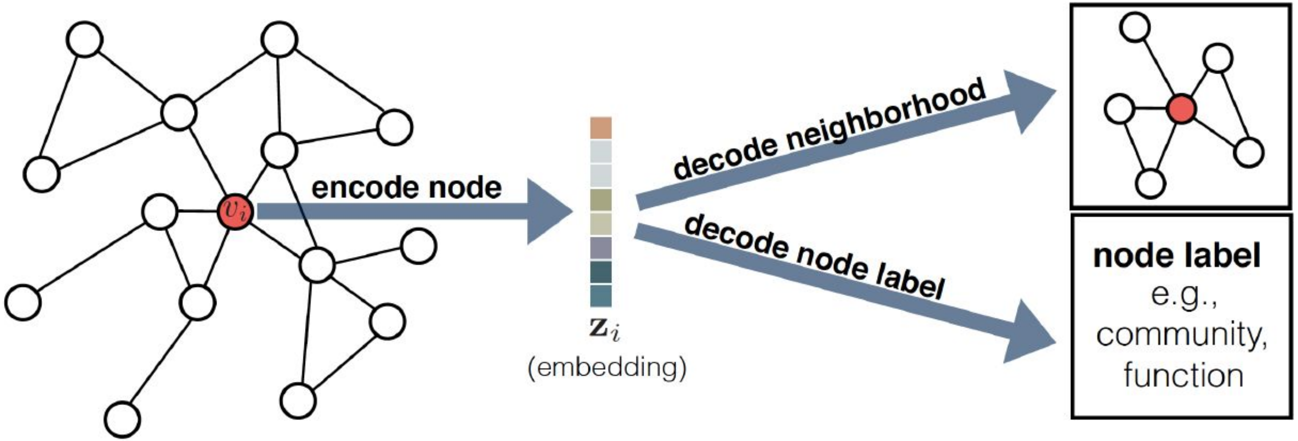


Fig. 1. A toy example of embedding a graph into 2D space with different granularities. $G_{\{1,2,3\}}$ denotes the substructure containing node v_1, v_2, v_3.

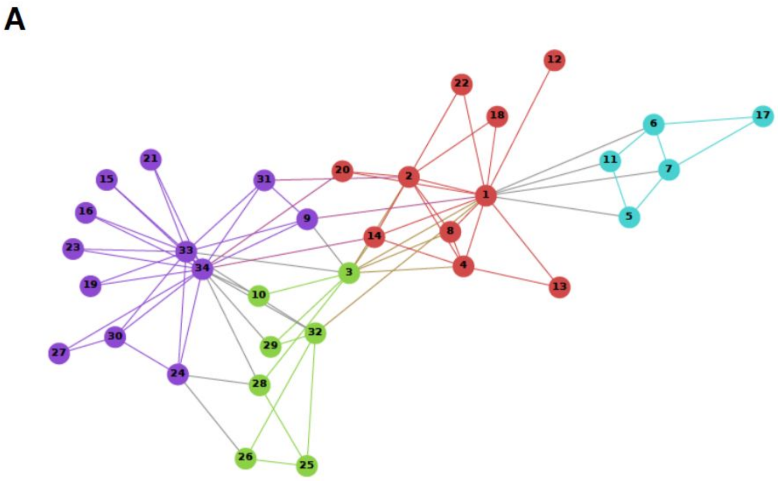
Source: H. Cai et al., A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications

Node embedding

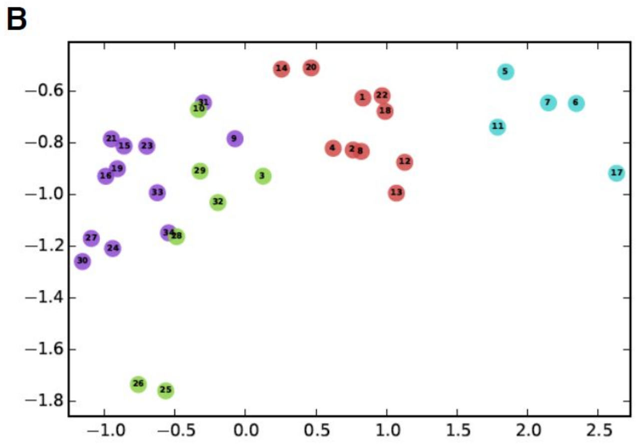


Node embedding

- “near” nodes should have a similar vector representation
- node proximity measure:
  - usually 1st or 2nd order node proximity



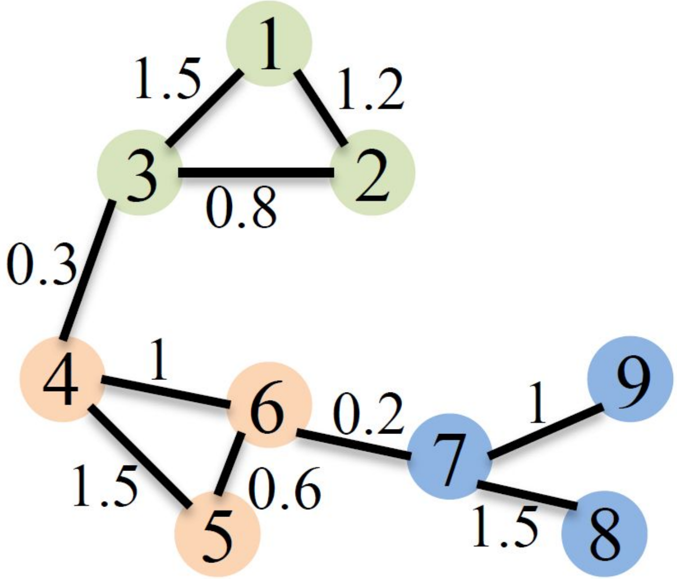
Zachary Karate Club social network



2D visualization of node embeddings

1st order node proximity

1st order neighborhood $s_{ij}^{(1)}$ of nodes v_i and v_j is the weight of the edge e_{ij} between those.



$$s_i^{(1)}=[s_{i1}^{(1)}, s_{i2}^{(1)},\ldots, s_{i|V|}^{(1)}]$$

$$s_1^{(1)}=[s_{11}^{(1)}, s_{12}^{(1)},\ldots, s_{19}^{(1)}]$$

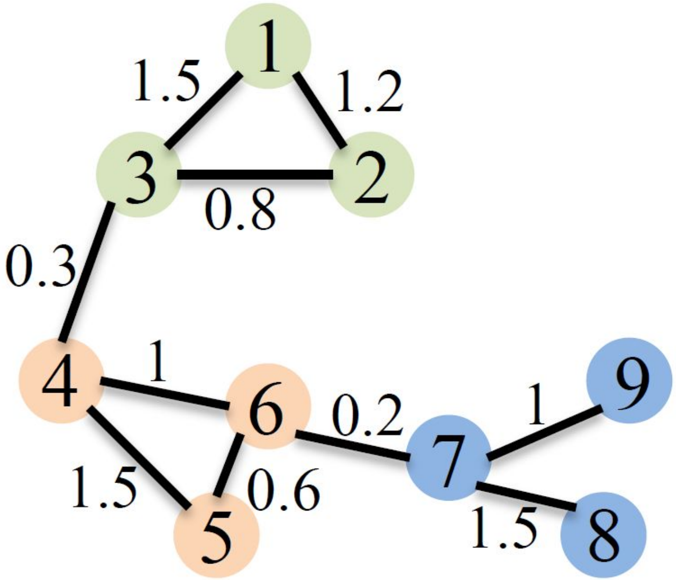
$$s_1^{(1)}=[0, 1.2, 1.5, 0, 0, 0, 0, 0, 0]$$

Source: H. Cai et al., A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications

2nd order node proximity

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}},$$

2nd order neighborhood $s_{ij}^{(2)}$ of nodes v_i a v_j is the similarity of their 1st order node neighborhoods: $s_i^{(1)}$ for node v_i and $s_j^{(1)}$ for node v_j.



$s_1^{(1)}$=[0, 1.2, 1.5, 0, 0, 0, 0, 0, 0]

$s_2^{(1)}$=[1.2, 0, 0.8, 0, 0, 0, 0, 0, 0]

$s_{12}^{(2)}$ =cosine($s_1^{(1)}$, $s_2^{(1)}$)=0.43

$s_{15}^{(2)}$ =cosine($s_1^{(1)}$, $s_5^{(1)}$)=0

0 => no common neighbours

Source: *H. Cai et al., A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications*

Edge embedding

Operator	Symbol	Definition
Average	$\boxplus$	$[f(u) \boxplus f(v)]_i = \frac{f_i(u)+f_i(v)}{2}$
Hadamard	$\boxdot$	$[f(u) \boxdot f(v)]_i = f_i(u) * f_i(v)$
Weighted-L1	$\Vert \cdot \Vert_{\bar{1}}$	$\Vert f(u) \cdot f(v) \Vert_{\bar{1}i} = \lvert f_i(u) - f_i(v) \rvert$
Weighted-L2	$\Vert \cdot \Vert_{\bar{2}}$	$\Vert f(u) \cdot f(v) \Vert_{\bar{2}i} = \lvert f_i(u) - f_i(v) \rvert^2$

u, v - nodes
f(u), f(v) - embedding vectors for nodes u i v

Edge embedding

LINK PREDICTION

<u>Prediction of missing edges</u>

- finding such edges in datasets

<u>Prediction of “probable” edges</u>

- recommender systems: friends, movies
- recommendation of potential scientific research topics

Edge embedding
KNOWLEDGE GRAPHS

- knowledge database
- edge: <h, r, t>, <head entity, relation, tail entity>
- directed graph
- finding missing entity/relation based on the other two

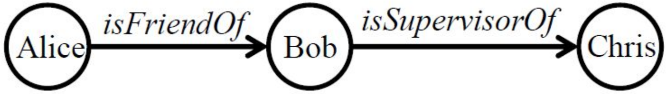


Fig. 3. A toy example of knowledge graph.

<Alice, isFriendOf, Bob>
<Bob, isSupervisorOf, Chris>

Source: H. Cai et al., A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications

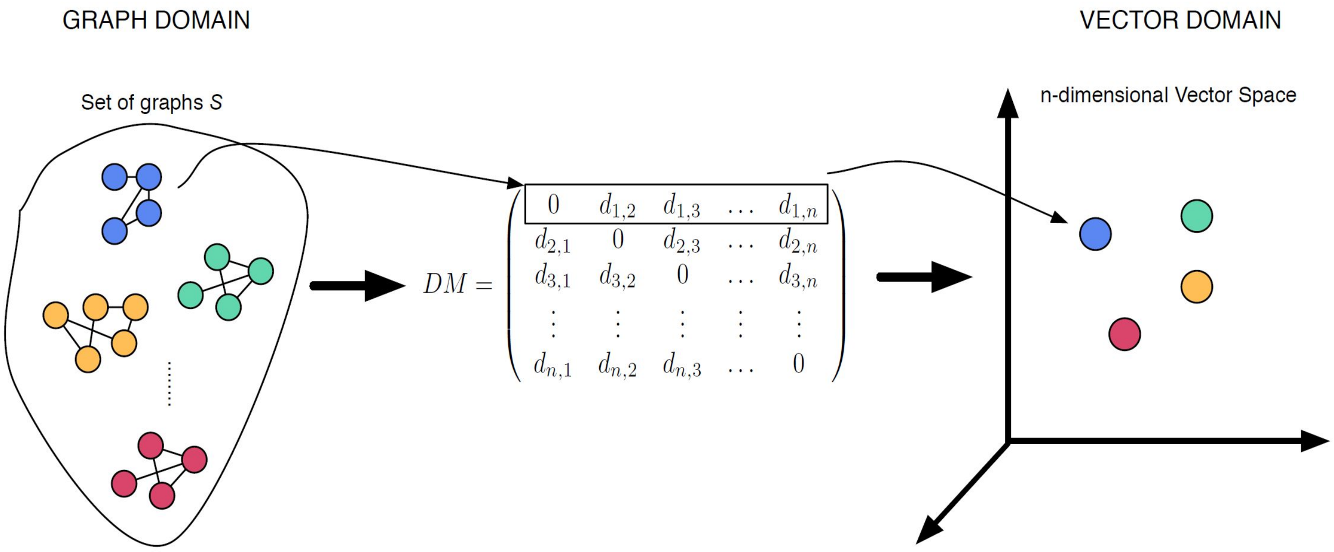
Subgraph / whole graph embedding

<u>Whole graph embedding</u>

comparing graph structures

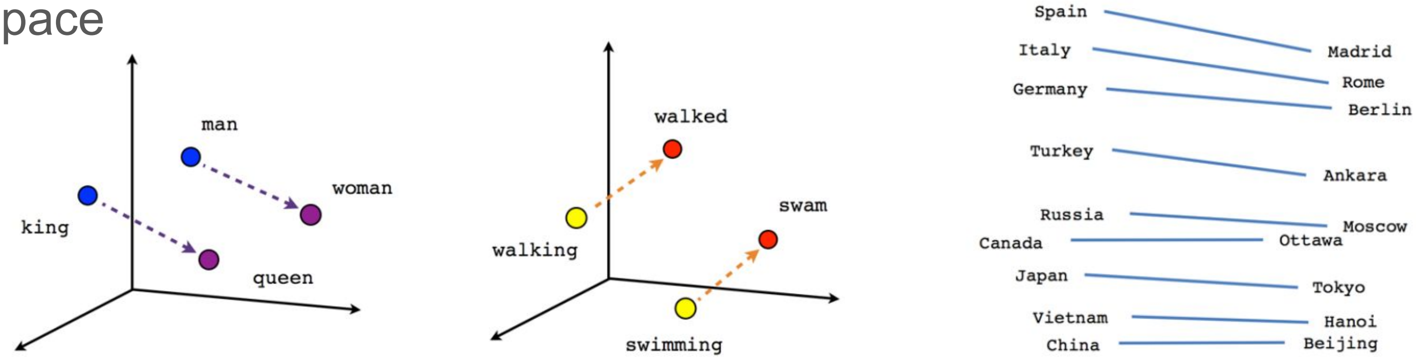
<u>Subgraph embedding</u>

comparing communities in graphs



Multi-Modal Bayesian Embeddings for Learning Social Knowledge Graphs

- learn **latent topics** that generate **word embeddings** and **network embeddings** simultaneously
- representation of social network users and knowledge concepts in a shared latent topic space

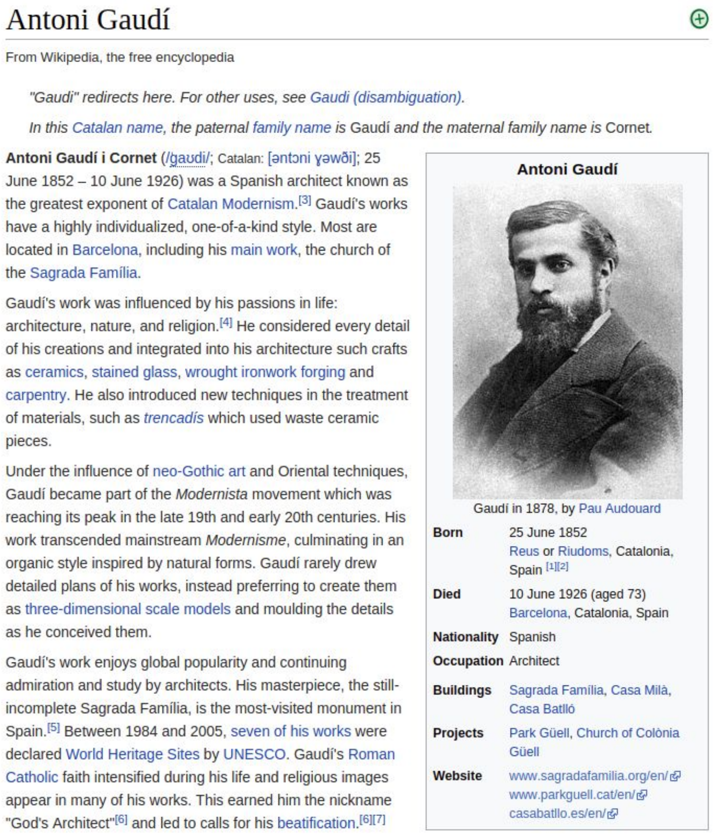


Male-Female | Verb tense | Country-Capital

Yang, Z., Tang, J., & Cohen, W. (n.d.). *Multi-Modal Bayesian Embeddings for Learning Social Knowledge Graphs*.

Input

- social network $G^r = (V^r, E^r)$
  - V^r set of users
  - E^r set of edges between the users
- knowledge base $G^k = (V^k, C)$
  - V^k set of knowledge concepts
  - C text associated with or facts between the concepts
- text posted by users of the social network D
  - Given a user $u \in V^r$, $d_u \in D$ denotes a document of all text posted by u (each user u has only one document)



Antoni Gaudí

From Wikipedia, the free encyclopedia

"Gaudi" redirects here. For other uses, see Gaudi (disambiguation).

In this Catalan name, the paternal family name is Gaudí and the maternal family name is Cornet.

Antoni Gaudí i Cornet (/ɡaʊdi/; Catalan: [əntɔni ɣəwði]; 25 June 1852 – 10 June 1926) was a Spanish architect known as the greatest exponent of Catalan Modernism.[3] Gaudí's works have a highly individualized, one-of-a-kind style. Most are located in Barcelona, including his main work, the church of the Sagrada Família.

Gaudí's work was influenced by his passions in life: architecture, nature, and religion.[4] He considered every detail of his creations and integrated into his architecture such crafts as ceramics, stained glass, wrought ironwork forging and carpentry. He also introduced new techniques in the treatment of materials, such as *trencadís* which used waste ceramic pieces.

Under the influence of neo-Gothic art and Oriental techniques, Gaudí became part of the *Modernista* movement which was reaching its peak in the late 19th and early 20th centuries. His work transcended mainstream *Modernisme*, culminating in an organic style inspired by natural forms. Gaudí rarely drew detailed plans of his works, instead preferring to create them as three-dimensional scale models and moulding the details as he conceived them.

Gaudí's work enjoys global popularity and continuing admiration and study by architects. His masterpiece, the still-incomplete Sagrada Família, is the most-visited monument in Spain.[5] Between 1984 and 2005, seven of his works were declared World Heritage Sites by UNESCO. Gaudí's Roman Catholic faith intensified during his life and religious images appear in many of his works. This earned him the nickname "God's Architect"[6] and led to calls for his beatification.[6][7]

Antoni Gaudí

Gaudí in 1878, by Pau Audouard

Born	25 June 1852 Reus or Riudoms, Catalonia, Spain [1][2]
Died	10 June 1926 (aged 73) Barcelona, Catalonia, Spain
Nationality	Spanish
Occupation	Architect
Buildings	Sagrada Família, Casa Milà, Casa Batlló
Projects	Park Güell, Church of Colònia Güell
Website	www.sagradafamilia.org/en/ www.parkguell.cat/en/ casabatllo.es/en/

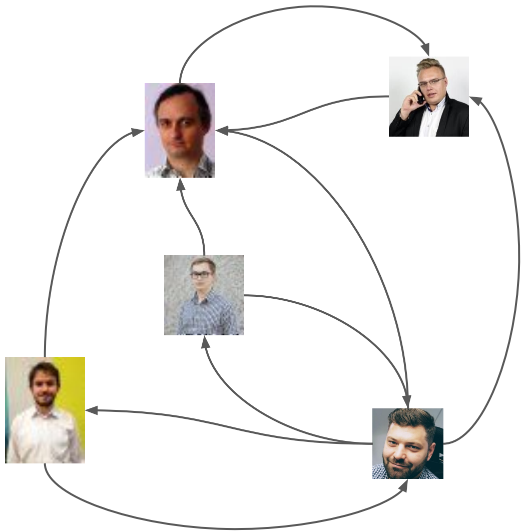
Output

- social knowledge graph $G = (V^r, V^k, P)$
- given a user $u \in V^r$, P_u is a ranked list of top-k knowledge concepts in V^k, where order indicates the relatedness to user u
- E.g. academic social network
  - top-k research interests of each researcher

1	Machine Learning
2	Embedding
3	Bayesian inference
4	Social Networks
5	Social Media

Two modalities

social network of users



knowledge concepts

Beta distribution

From Wikipedia, the free encyclopedia

Not to be confused with Beta function.

In probability theory and statistics, the **beta distribution** is a family of continuous probability distributions defined on the interval [0, 1] parametrized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution. It is a special case of the Dirichlet distribution.

The beta distribution has been applied to model the behavior of random variables limited to intervals of finite length in a wide variety of disciplines.

In Bayesian inference, the beta distribution is the conjugate prior probability distribution for the Bernoulli, binomial, negative binomial and geometric distributions. For example, the beta distribution can be used in Bayesian analysis to describe initial knowledge concerning probability of success such as the probability that a space vehicle will successfully complete a specified mission. The beta distribution is a suitable model for the random behavior of percentages and proportions.

The usual formulation of the beta distribution is also known as the **beta distribution of the first kind**, whereas *beta distribution of the second kind* is an alternative name for the beta prime distribution.

Contents [hide]

More about input

- pretrained knowledge concept embeddings
  - encode information from the knowledge base G^k
  - *e.g. skip-gram model [Mikolov et al., 2013]*
- pretrained user embeddings as input
  - social network G^r
  - *e.g. DeepWalk [Perozzi et al., 2014]*

The model

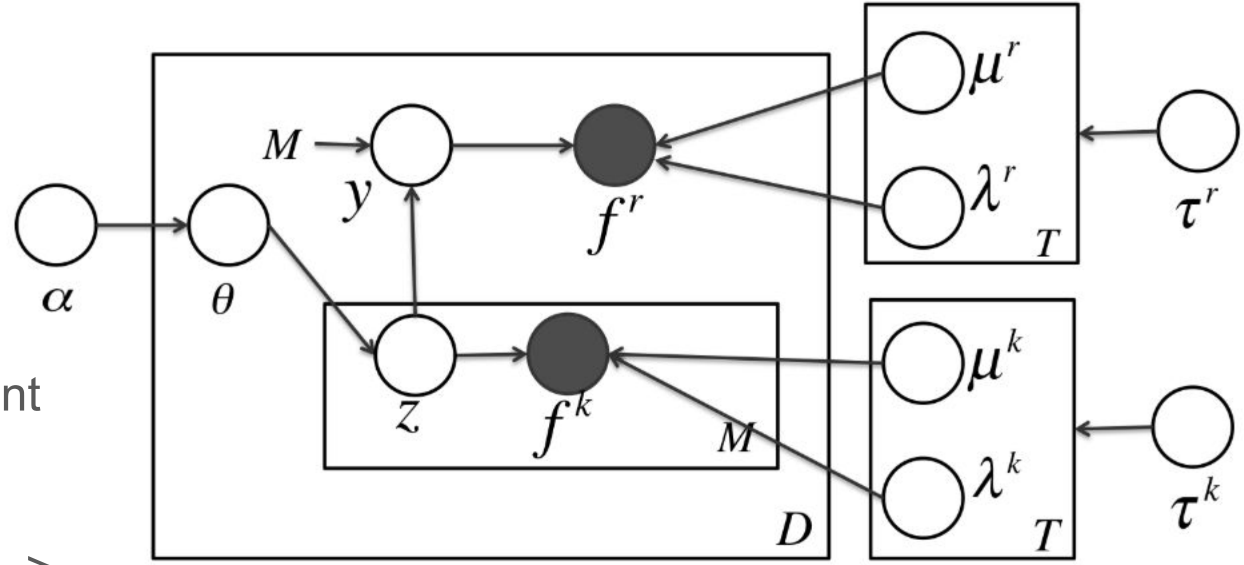
D - documents

Z - topics

M - concepts in a document

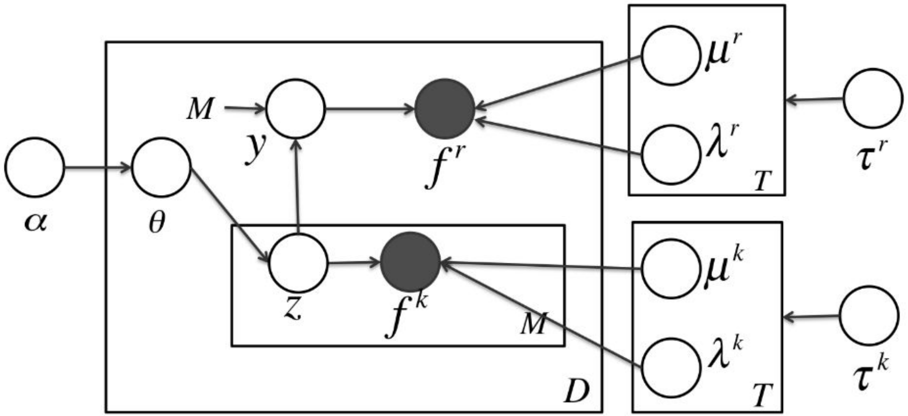
f^r - embeddings of users drawn from Normal(μ,1/λ) -> from NormalGamma(τ^r)

f^k_{um} - embedding of *m*-concept in document d_u



θ_u - multinomial topic distribution of document d_u (or user *u*)
z_{um} - topic of the *m*-th knowledge concept in document d_u
y_u - topics of user *u*

Generative process



1. For each topic t, and for each dimension
   (a) Draw μ_t^r, λ_t^r from NormalGamma(τ^r)
   (b) Draw μ_t^k, λ_t^k from NormalGamma(τ^k)
2. For each user u
   (a) Draw a multinomial distribution θ from Dir(α)
   (b) For each knowledge concept w in d_u
      i. Draw a topic z from Multi(θ)
      ii. For each dimension of the embedding of w, draw f^k from $\mathcal{N}(\mu_z^k, \lambda_z^k)$
   (c) Draw a topic y uniformly from all z's in d_u
   (d) For each dimension of the embedding of user u, draw f^r from $\mathcal{N}(\mu_y^r, \lambda_y^r)$

Inference

- collapsed Gibbs sampling [Griffiths, 2002]
- extension of Gaussian LDA [Das et al., 2015] that updates the embeddings during inference

Experiments

Data for embedding calculations:

- AMiner - co-authorships between researchers
  - documents of authors
- Wikipedia The Free Encyclopedia - knowledge base of articles
  - (Wikipedia corpus to learn the knowledge concept embeddings)

Evaluation:

Ranking of concepts from:

- homepage
- linkedin
- crowdsourcing

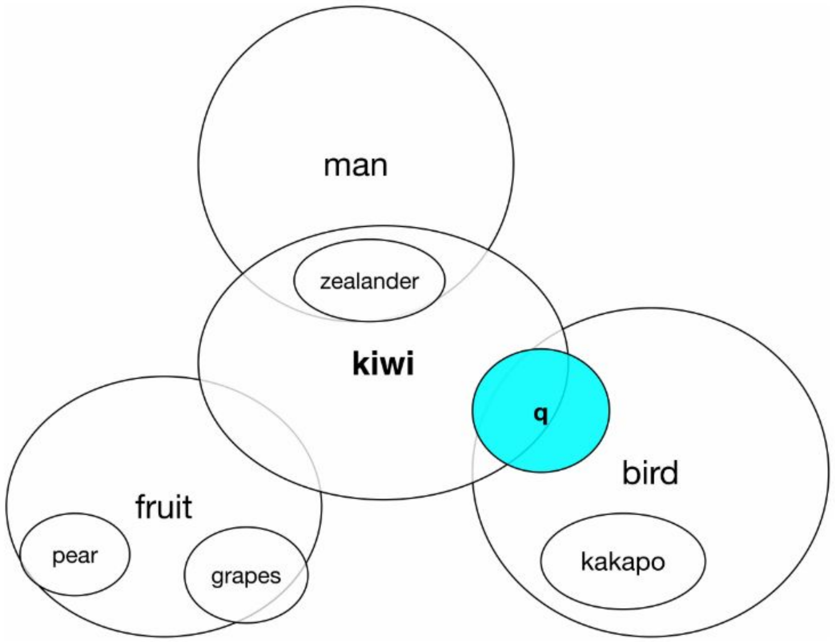
1	Machine Learning
2	Embedding
3	Bayesian inference
4	Social Networks
5	Social Media

Topic #1	Topic #2	Topic #3
GenVector		
query expansion	image processing	hepatocellular carcinoma
concept mining	face recognition	gastric cancer
language modeling	feature extraction	acute lymphoblastic leukemia
information extraction	computer vision	renal cell carcinoma
knowledge extraction	image segmentation	glioblastoma multiforme
entity linking	image analysis	acute myeloid leukemia
language models	feature detection	peripheral blood
named entity recognition	digital image processing	malignant melanoma
document clustering	machine learning algorithms	hepatitis c virus
latent semantic indexing	machine vision	squamous cell carcinoma
Thorsten Joachims	Anil K. Jain	Keizo Sugimachi
Jian Pei	Thomas S. Huang	Setsuo Hirohashi
Christopher D. Manning	Peter N. Belhumeur	Masatoshi Makuuchi
Raymond J. Mooney	Azriel Rosenfeld	Morito Monden
Charu C. Aggarwal	Josef Kittler	Yoshio Yamaoka
William W. Cohen	Shuicheng Yan	Kunio Okuda
Eugene Charniak	David Zhang	Yasuni Nakanuma
Kamal Nigam	Xiaoou Tang	Kendo Kiyosawa
Susan T. Dumais	Roberto Cipolla	Masazumi Tsuneyoshi
T. K. Landauer	David A. Forsyth	Satoru Todo

Embedding Words as Distributions with a Bayesian Skip-gram Model

Idea:

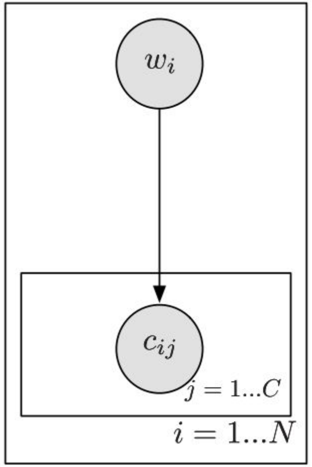
- replace point word embedding with distribution



Havrylov, Serhii, and Ivan Titov. "Embedding Words as Distributions with a Bayesian Skip-gram Model." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

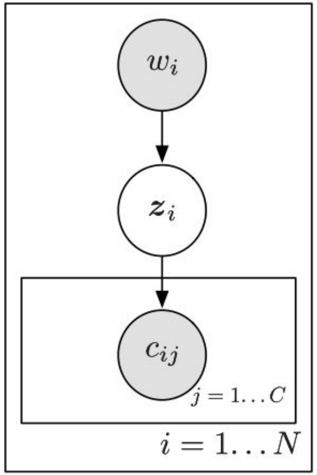
Models

skip-gram



- For each data point $i = 1...N$
  - $\star$ For each context word $j = 1...C$
    - $\cdot$ Draw a context word $c_{ij} \sim p(c|w_i)$

Bayesian skip-gram



- For each data point $i = 1...N$
  - $\star$ Draw a latent vector $\boldsymbol{z}_i \sim p(\boldsymbol{z}|w_i)$
  - $\star$ For each context word $j = 1...C$
    - $\cdot$ Draw a context word $c_{ij} \sim p(c|\boldsymbol{z}_i, w_i)$

Examples

word 1	word 2	KL	cosine sim.
dog	cat	15.47	0.71
dog	pet	18.52	0.70
dog	hound	21.20	0.64
dog	animal	27.69	0.52
cappuccino	espresso	12.59	0.76
cappuccino	latte	13.39	0.7
cappuccino	coffee	22.54	0.69
cappuccino	drink	30.81	0.54
microsoft	windows	24.41	0.65
microsoft	google	24.44	0.60
microsoft	corporation	39.40	0.29
microsoft	company	46.05	0.19