MultiVERSE: a multiplex and multiplex-heterogeneous network embedding approach

Léo Pio-Lopez, Alberto Valdeolivas, Laurent Tichit, Élisabeth Remy & Anaïs Baudot

Published in 2021 in scientific reports

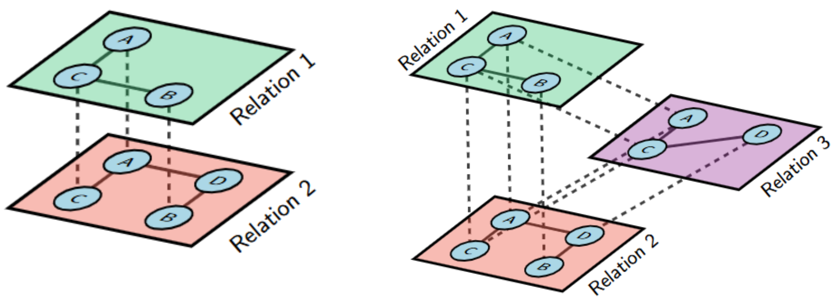
11.03.2025

Clemence Reda

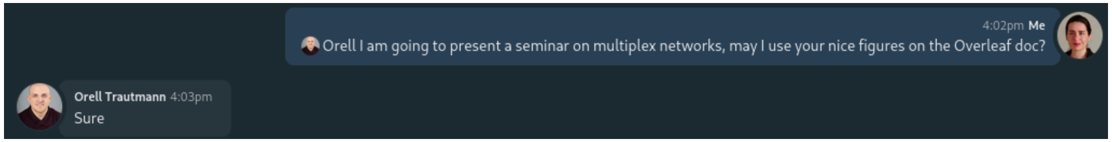
Background

1. Multiplex networks
2. Multiplex-heterogeneous networks
3. Network embedding learning
4. Downstream applications

Multiplex network a graph with N nodes and L layers. Only interlayer edges between replicates are allowed.



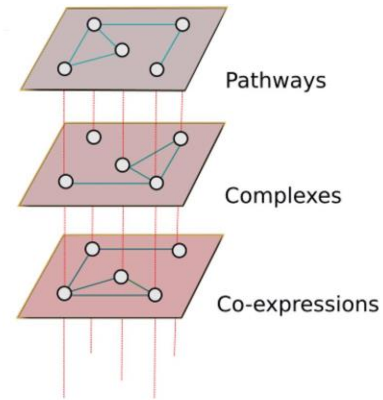
Courtesy of Orell Trautmann (SBI Rostock).



Replicate copy of a node from another layer
Edges can be directed, weighted
⚠ #edge types ≠ #layers in general
eg., Twitter **[1]**, Leukemia **[2]**, PPIs

Goal Encode complex (*eg*., regulatory) interactions across relations/layers in a systematic fashion

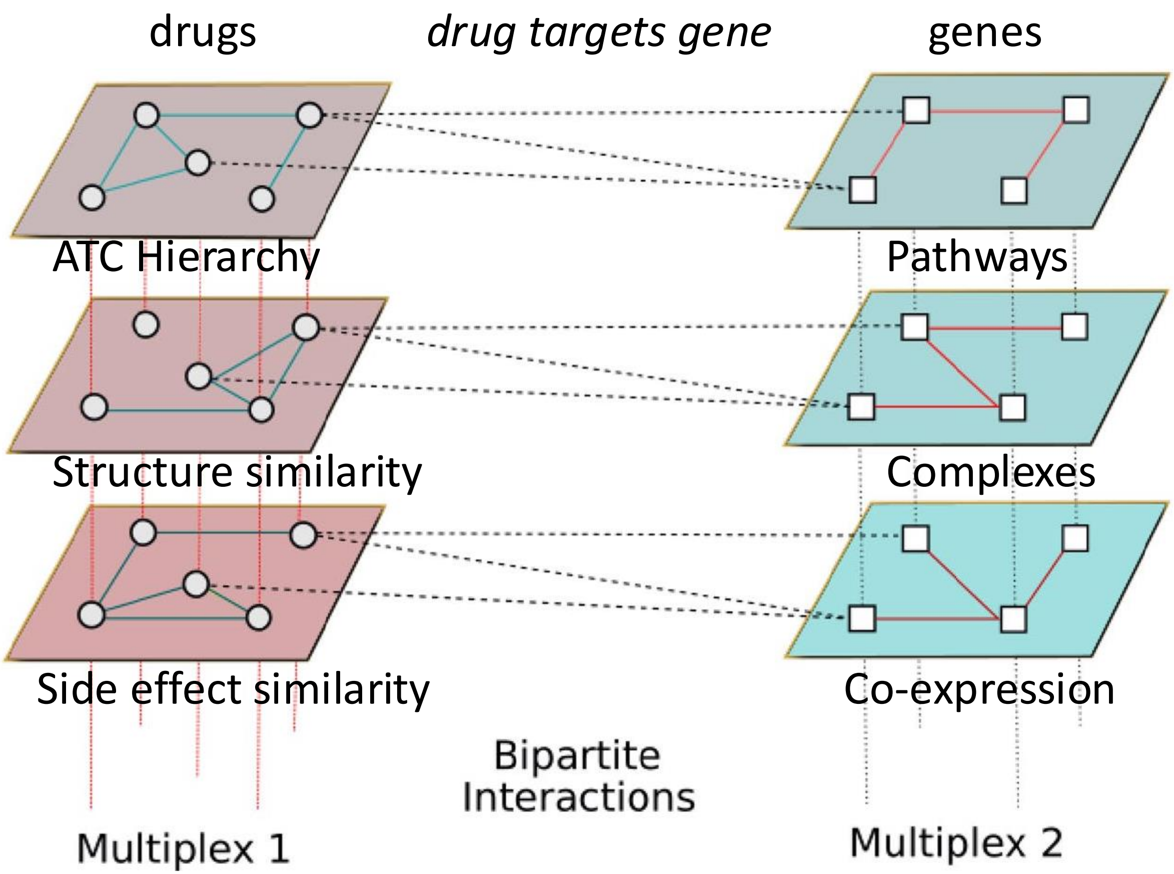
Size #nodes ranges from 10 (Vickers) to 450k (Twitter)
#edges: from 1k to 15M #layers: from 3 to 18 (as far as I saw)



Typical biological multiplex network (from the paper).

[1] https://github.com/galadrielbriere/ApplicationsMultiXrank/tree/main/Leukemia
[2] https://snap.stanford.edu/data/higgs-twitter.html

Multiplex-heterogeneous network a set of multiplex networks connected by bipartite interactions.



Multiplex-heterogeneous network (adapted from the MultiVERSE paper).

Connect different sets of nodes (and get a network of networks).

Bipartite interactions are preserved across layers.

❓ Well, could we just aggregate everything in a single layer and reannotate edges?

✓Having different layers/topologies has an impact on the exploration process (eg., multi-edge vs multiplex networks)

Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods ← In the MultiVERSE paper
eg., node2vec, DeepWalk

Local methods (the more you know)
eg., translational models (TransE, MurE, ...)

Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods ⬅ In the MultiVERSE paper
eg., node2vec, DeepWalk

Local methods (the more you know)
eg., translational models (TransE, MurE, ...)

Johnson–Lindenstrauss lemma

Given $\varepsilon \in (0,1)$, m points $\boxed{x_1, \dots, x_m}$ in $\mathbb{R}^N$, and an integer $n > 8\ln(m)/\varepsilon^2$, there exists a linear map $f : \mathbb{R}^N \to \mathbb{R}^n$ such that rows of A (adjacency matrix of a network)

$$(1-\varepsilon)\|x_i - x_j\|^2 \leq \|f(x_i) - f(x_j)\|^2 \leq (1+\varepsilon)\|x_i - x_j\|^2$$

embeddings

for all $i, j \in [\![1, m]\!]$.

Credit to Felix Lefebvre, Inria SODA.

TL;DR If the embedding dimension is not too small, we can hope for a good representation

Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

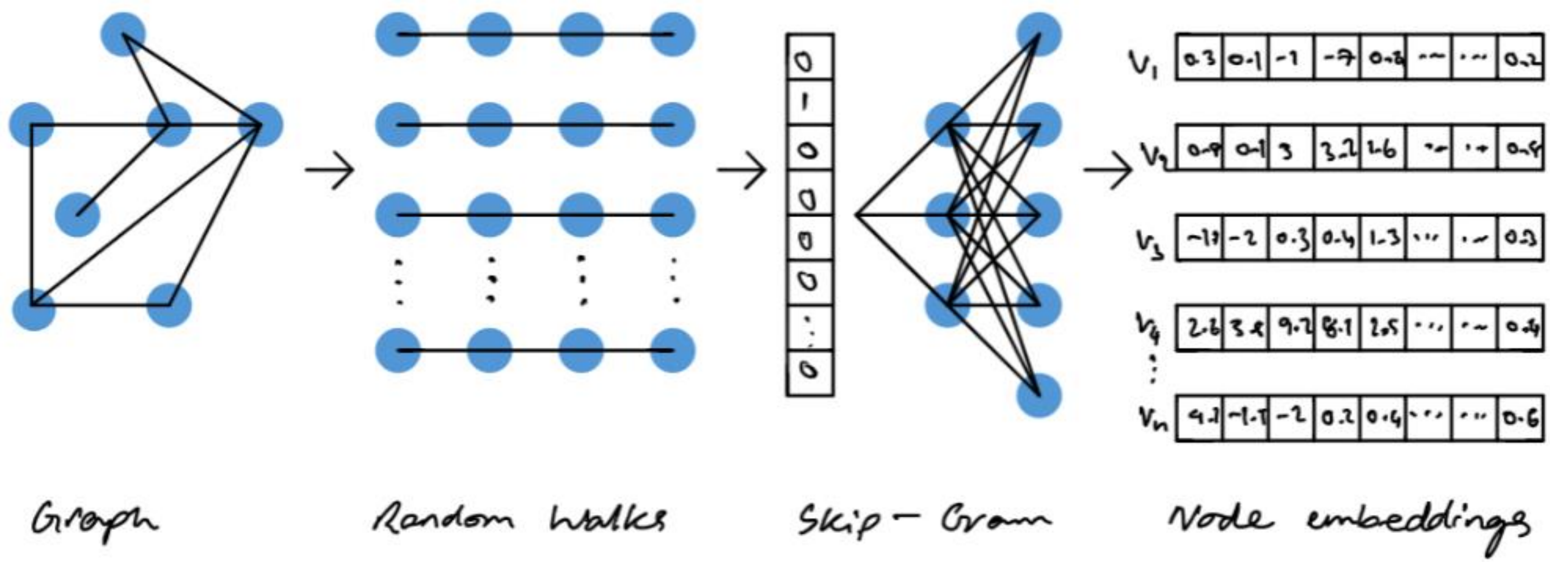
Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods — In the MultiVERSE paper

eg., node2vec, DeepWalk

Local methods (the more you know)

eg., translational models (TransE, MurE, ...)



Credit to Vatsal, "Node2Vec Explained", Medium.

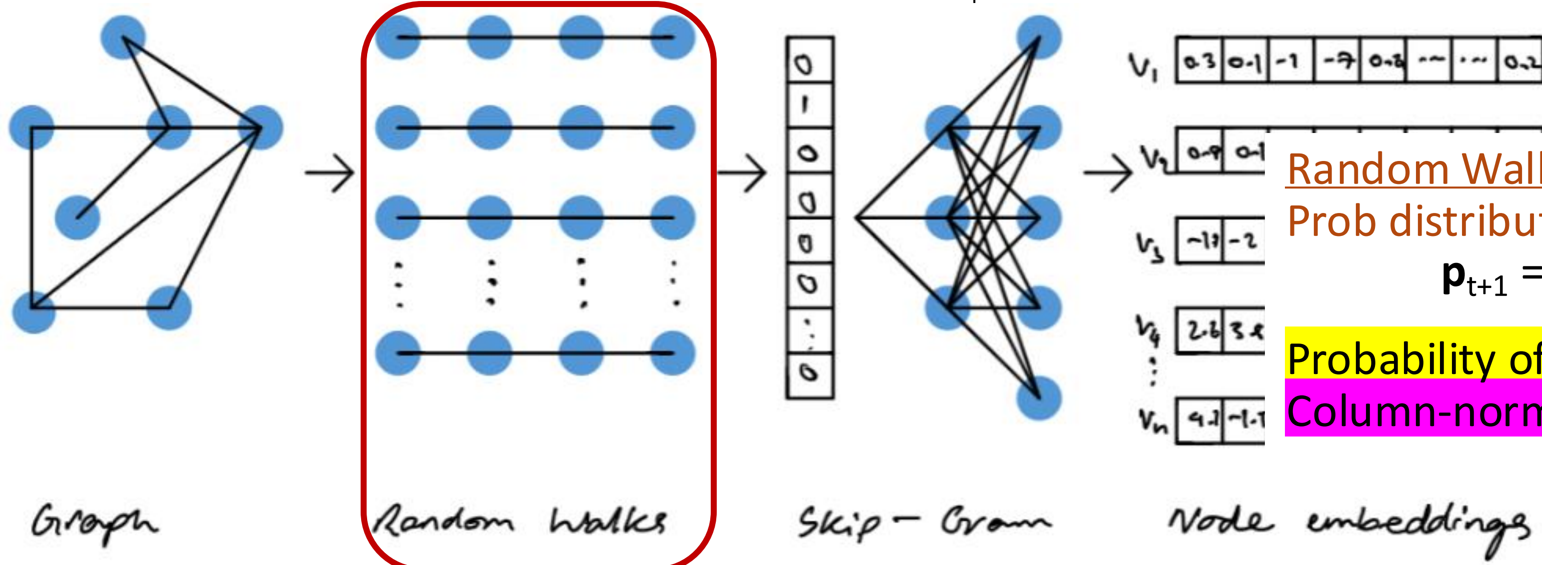
Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods
eg., node2vec, DeepWalk

In the MultiVERSE paper

Local methods (the more you know)
eg., translational models (TransE, MurE, ...)



Random Walk with Restart (RWR)
Prob distribution of nodes varies as

$$\mathbf{p}_{t+1} = (1-r)M\mathbf{p}_t + r\mathbf{p}_0$$

Probability of restart (t=0)
Column-normalized adjacency mat

Credit to Vatsal, "Node2Vec Explained", Medium.

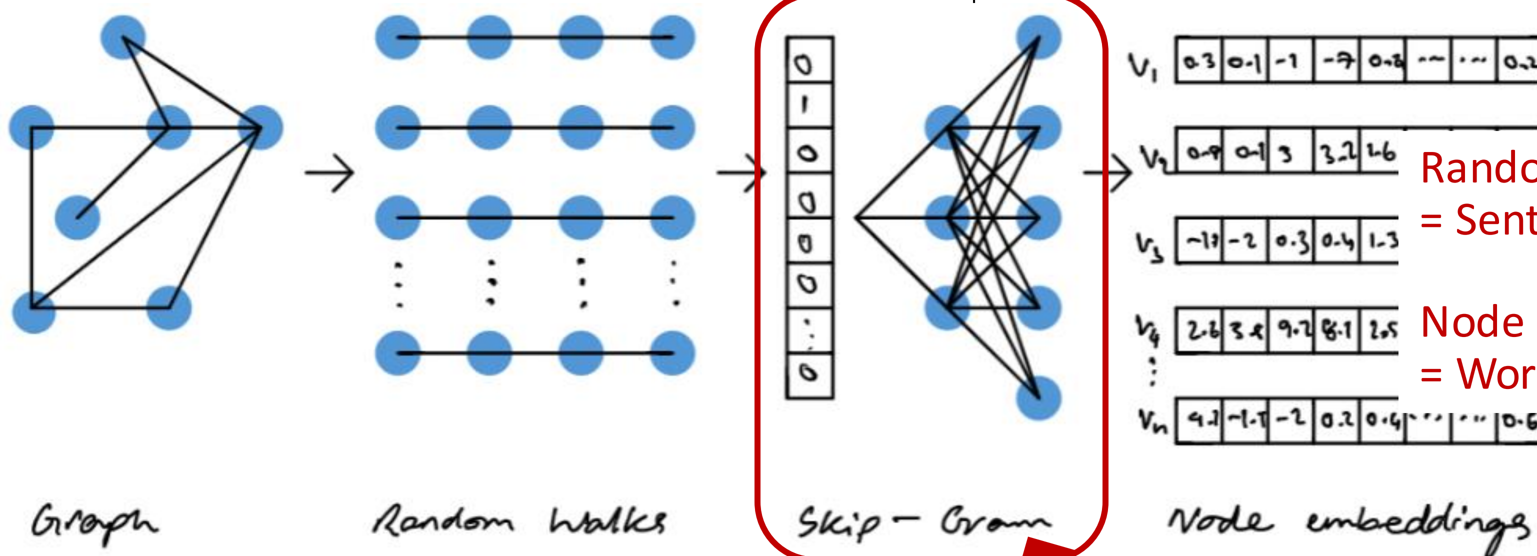
Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods
eg., node2vec, DeepWalk

In the MultiVERSE paper

Local methods (the more you know)
eg., translational models (TransE, MurE, ...)



Random walk = Sentence

Node = Word

Credit to Vatsal, "Node2Vec Explained", Medium.

(the one in word2vec)

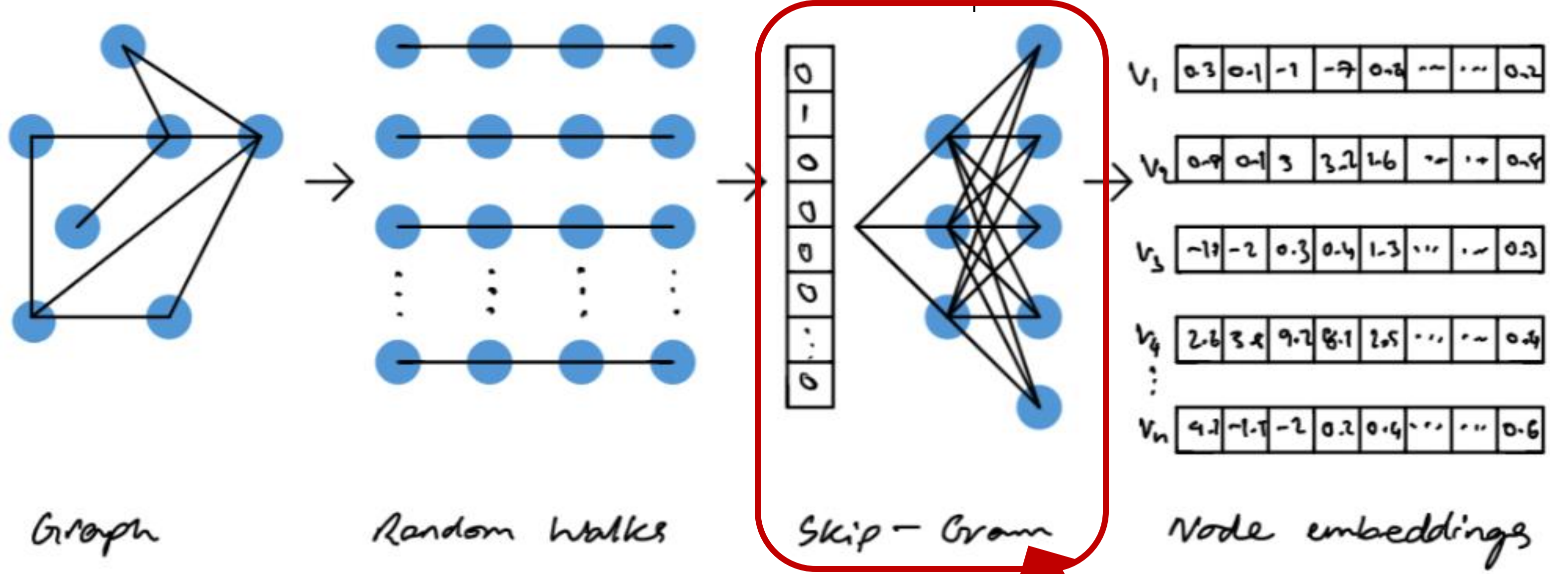
Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods
eg., node2vec, DeepWalk

In the MultiVERSE paper

Local methods (the more you know)
eg., translational models (TransE, MurE, ...)



"word" embeddings optimized for walking through neighbor nodes given a target node

Credit to Vatsal, "Node2Vec Explained", Medium.

(the one in word2vec)

Network embedding learning learning a low-dimensional representation for each node (≈ automatic feature crafting).

Why? A network-agnostic way to exploit node connections.
Meaning Relative positions of embeddings bear information

Global methods ← In the MultiVERSE paper

eg., node2vec, DeepWalk

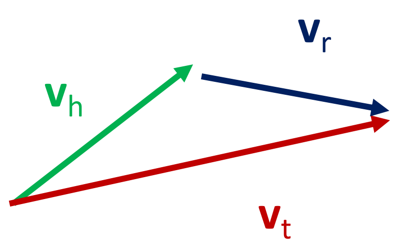
You can also use Graph Convolutional Networks for this:
"Graph Neural Networks Series | Part 3 | Node embedding"
by Omar Hussein on Medium

Local methods (the more you know)

eg., translational models (TransE, MurE, ...)

Optimize node embeddings to predict the presence of edges in the network

edge (h: node, r: edge type, t: node)



Find $\mathbf{v}_h$ $\mathbf{v}_r$ $\mathbf{v}_t$ that maximize
$-|\mathbf{v}_h + \mathbf{v}_r - \mathbf{v}_t|$

Content of the paper

"[...] MultiVERSE, a fast, scalable and versatile embedding approach to learn node embeddings on **multiplex and multiplex-heterogeneous networks**"

Specific issues:

- Scalability
- Structure of multiplex-heterogeneous networks

(replicate nodes & ≠ node types)

Prior approaches on multiplex networks: essentially node2vec adaptations

Content of the paper

"[...] MultiVERSE, a fast, scalable and versatile embedding approach to learn node embeddings on **multiplex and multiplex-heterogeneous networks**"

1. The VERSE framework (prior work on similarity-based embeddings)
2. The MultiVERSE approach (this paper)
3. Experimental results

VERSE focuses embedding learning on **similarity metrics** instead of node neighborhoods (a la node2vec).

Objective Minimize the Kullback-Leibler divergence $\Sigma_{\{i \text{ node}\}} \mathrm{KL}(\underbrace{\text{simG}(i, .)}_{\text{Target (normalized) similarity vector}} , \underbrace{\text{simE}(i, .)}_{\text{Reconstructed similarity vector (with embeddings)}})$

$\text{simE}(i, j) \approx \mathbf{v}_i^T \mathbf{v}_j$

VERSE focuses embedding learning on **similarity metrics** instead of node neighborhoods (a la node2vec).

Objective Minimize the Kullback-Leibler divergence $\sum_{\{i \text{ node}\}} \mathrm{KL}(\underbrace{\mathrm{simG}(i, .)}_{\text{Target (normalized) similarity vector}}, \underbrace{\mathrm{simE}(i, .)}_{\text{Reconstructed similarity vector (with embeddings)}})$

$\mathrm{simE}(i, j) \approx v_i^T v_j$

No need to compute the full normalization on similarities

Use **Noise Contrastive Estimation**

Algorithm 1 VERSE

```
function VERSE(G, sim_G, d)
    W ← N(0, d^{-1})                    ▷ With W ∈ R^{|V|×d}
    repeat
        u ~ P                           ▷ Sample a node
        v ~ sim_G(u)                    ▷ Sample positive example
        W_u, W_v ← UPDATE(u, v, 1)
        for i ← 1 ... s do
            ṽ ~ Q(u)                    ▷ Sample negative example
            W_u, W_ṽ ← UPDATE(u, ṽ, 0)
    until converged
    return W
function UPDATE(u, v, D)                ▷ Logistic gradient update
    g ← (D − σ(W_u · W_v)) * λ
    W_u ← g * W_v
    W_v ← g * W_u
```

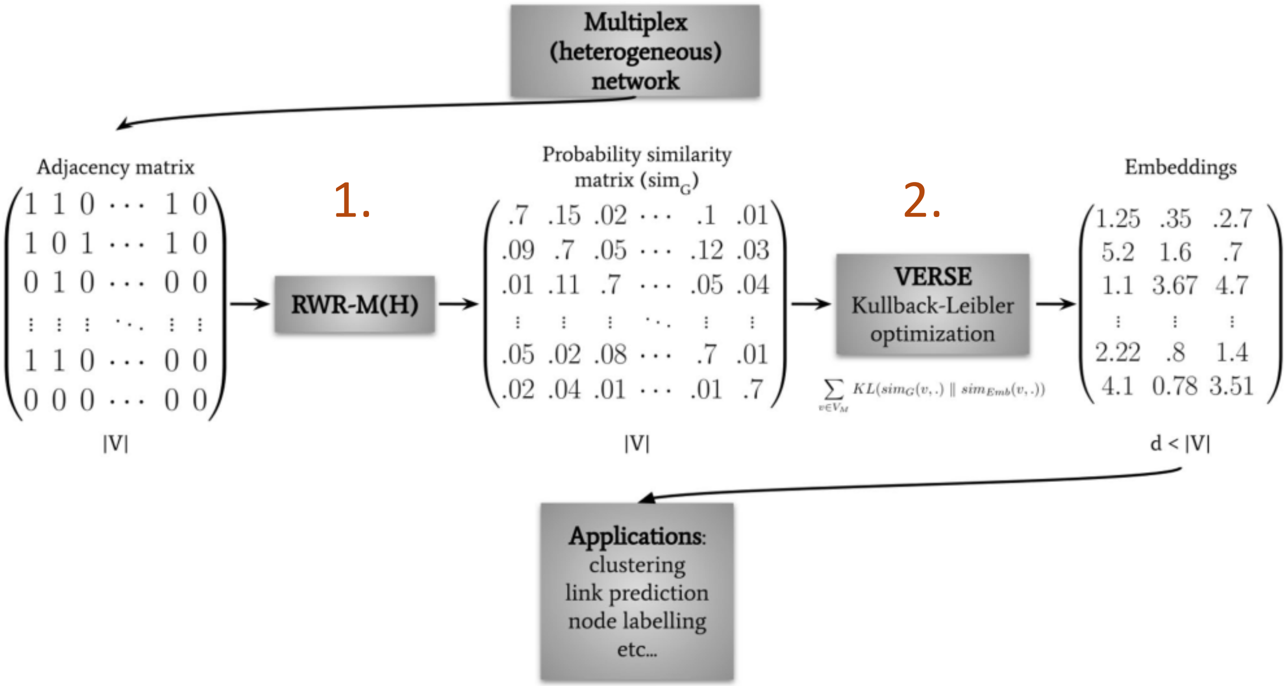
Approximate the probabilities for each node by discriminating with a logistic regressor

- nodes sampled by simG
- nodes sampled by a noise distribution Q

(in practice, at random)

From the VERSE paper.

MultiVERSE is an extension of VERSE to multiplex-het networks, with applications in biology and healthcare.

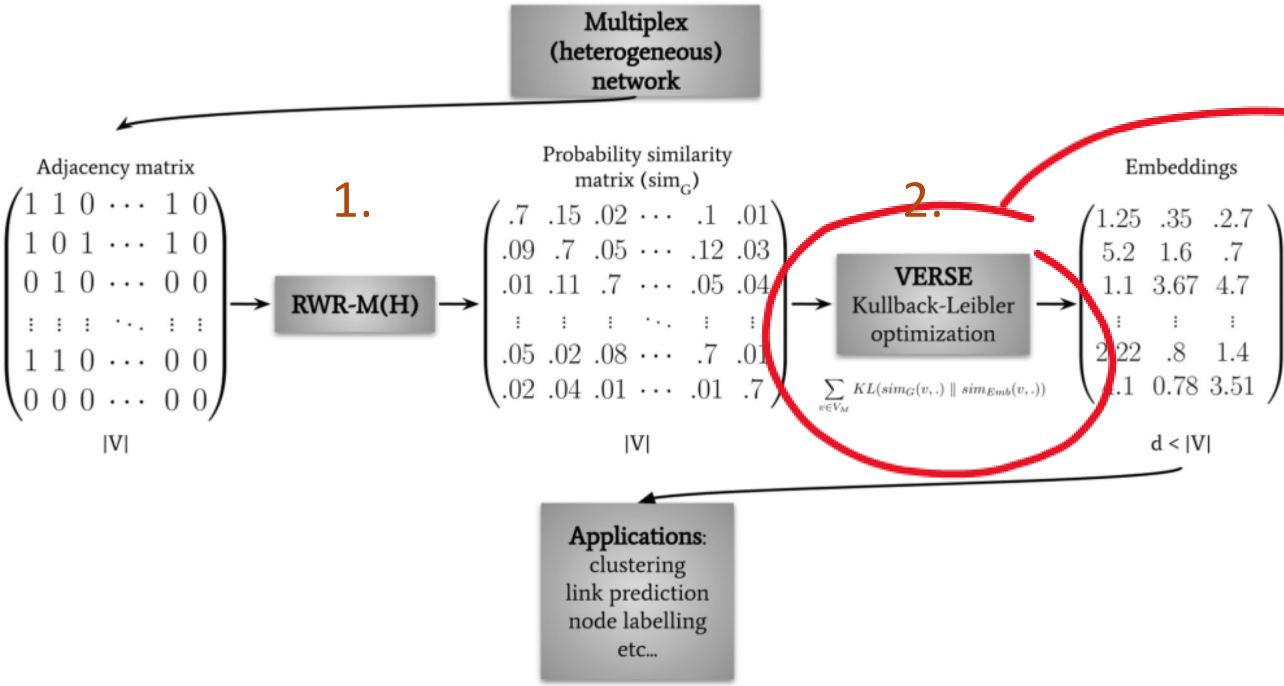


From the MultiVERSE paper.

1. Learn node similarities with a random walk specific to multiplex(-het) networks

2. Learn embeddings with VERSE

MultiVERSE is an extension of VERSE to multiplex-het networks, with applications in biology and healthcare.



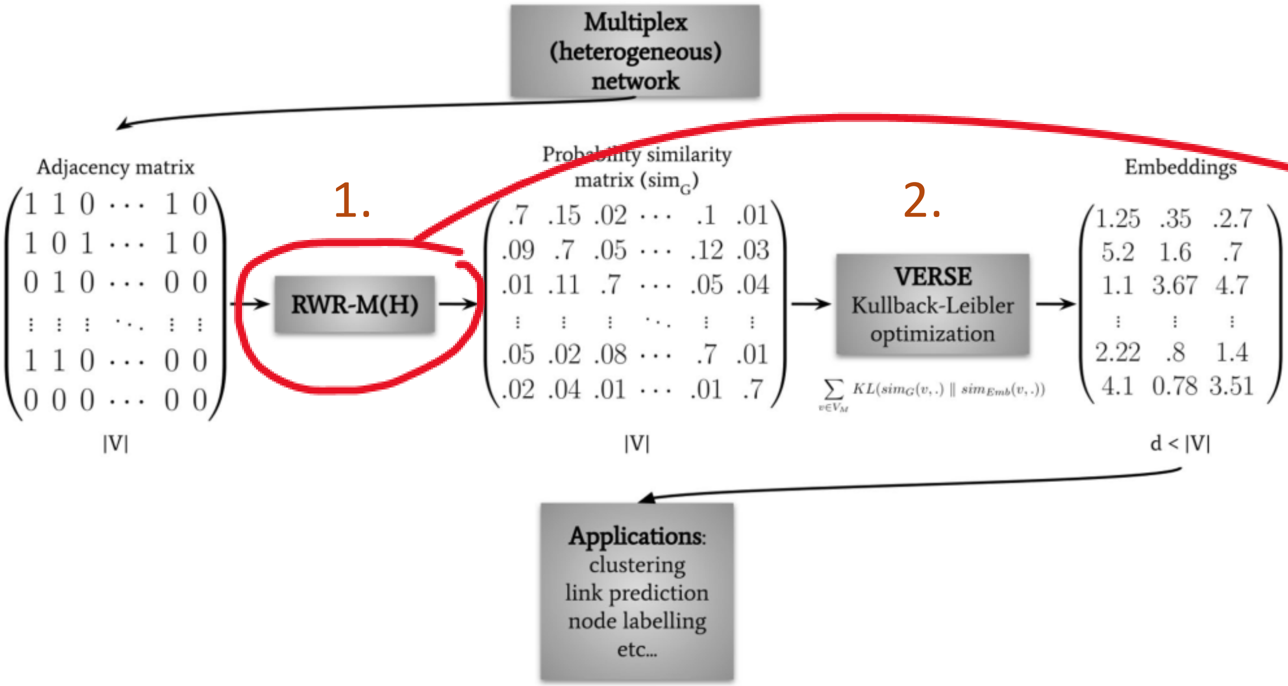
That was described in the previous slides

From the MultiVERSE paper.

1. Learn node similarities with a random walk specific to multiplex(-het) networks

2. **Learn embeddings with VERSE**

MultiVERSE is an extension of VERSE to multiplex-het networks, with applications in biology and healthcare.



From the MultiVERSE paper.

<u>RWR on multiplex nets (RWR-M)</u>

Prob distribution of nodes x layers varies as

$$\mathbf{p}_{t+1} = (1-r)M\mathbf{p}_t + r\ (\boldsymbol{\tau}\ .\mathbf{p}_0)$$

Probability of restart (t=0)

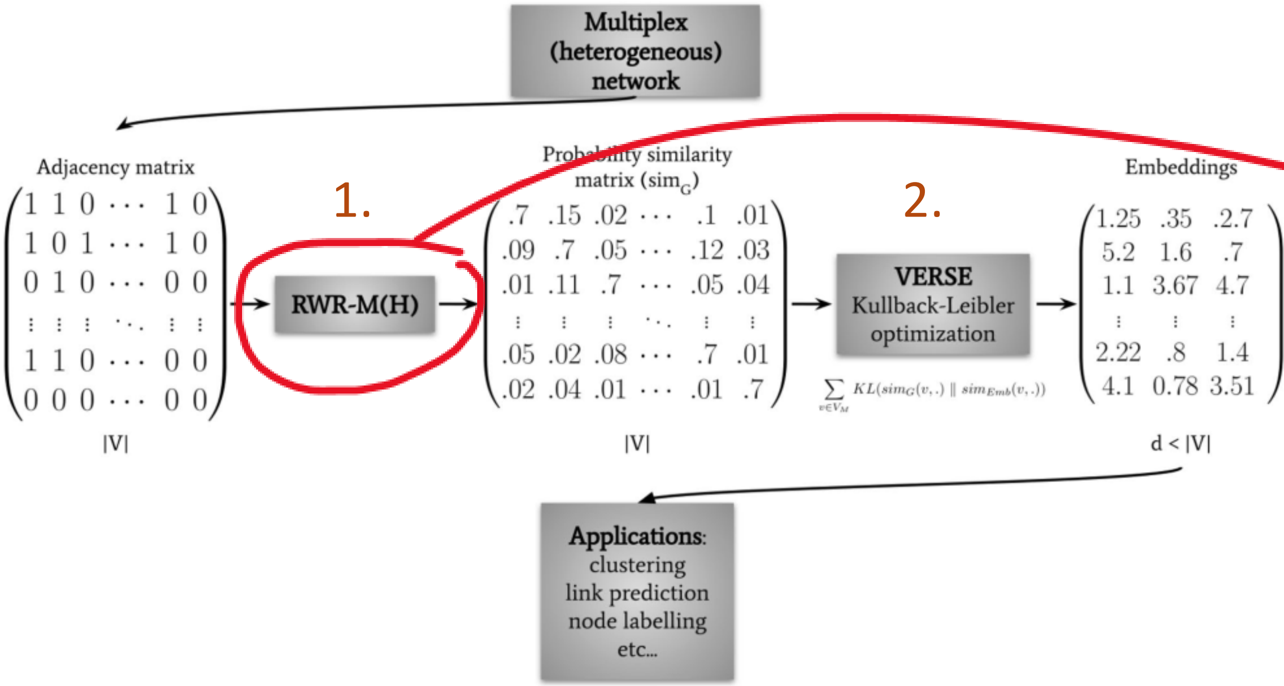
Prob of restart *in each layer*

Column-normalized adjacency mat defined by blocks

- <u>Intralayer:</u> adjacency matrix on this layer multiplied by 1-δ
- <u>Interlayer:</u> δ/(#layers - 1)

1. **Learn node similarities with a random walk specific to multiplex(-het) networks**
2. Learn embeddings with VERSE

MultiVERSE is an extension of VERSE to multiplex-het networks, with applications in biology and healthcare.



From the MultiVERSE paper.

<u>RWR on multiplex nets (RWR-M)</u>
Prob distribution of nodes x layers varies as

$$\mathbf{p}_{t+1} = (1-r)M\mathbf{p}_t + r\,(\boldsymbol{\tau}\,.\mathbf{p}_0)$$

Probability of restart (t=0)
Prob of restart *in each layer*
Column-normalized adjacency mat defined by blocks

- <u>Intralayer:</u> adjacency matrix on this layer multiplied by 1-δ
- <u>Interlayer:</u> δ/(#layers - 1)

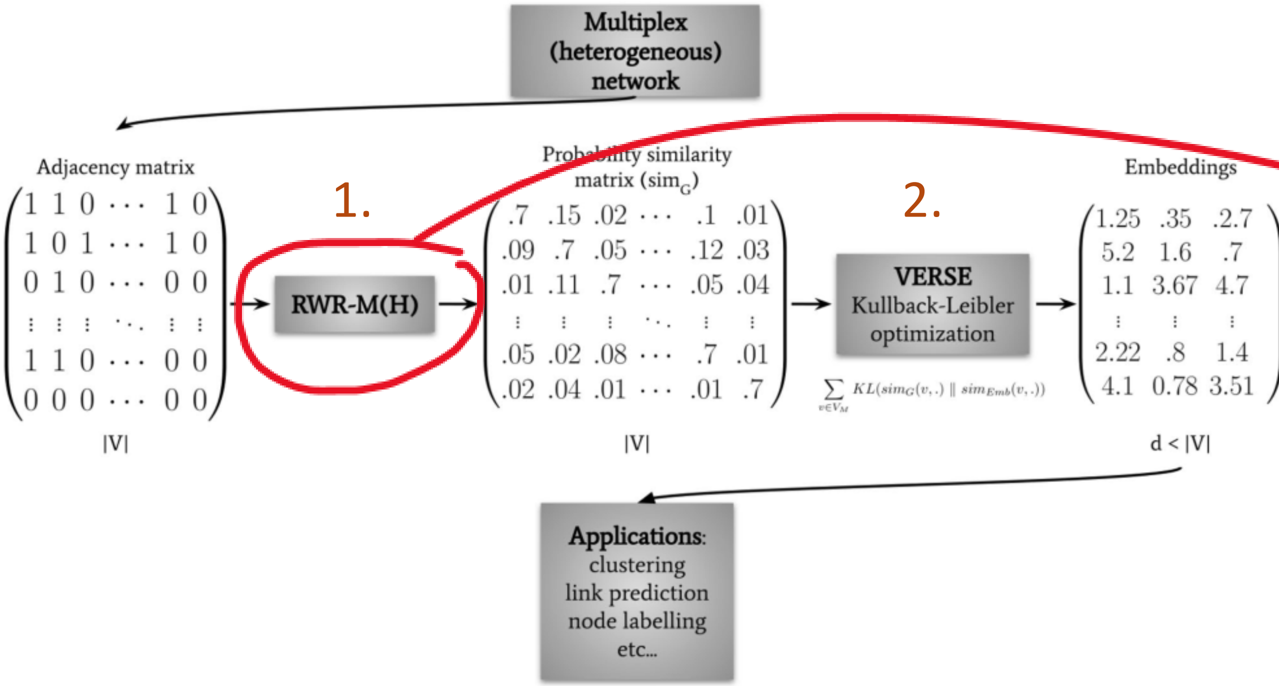
Prob of switching layers

Apply RWR-M #nodes times, stop at steady state
➡ #nodes x #nodes matrix for simG(.,.)

δ=0.5 r=0.7 **τ** uniform

1. **Learn node similarities with a random walk specific to multiplex(-het) networks**
2. Learn embeddings with VERSE

MultiVERSE is an extension of VERSE to multiplex-het networks, with applications in biology and healthcare.



From the MultiVERSE paper.

<u>RWR on multiplex-het nets (RWR-MH)</u>

N_1, N_2 are multiplex nets with L layers

Prob distribution of nodes x layers

$$\underline{\mathbf{p}}_{t+1} = (1-r)\mu\underline{p}_t + r\ (\underline{\boldsymbol{\tau}}\,.\underline{\mathbf{p}}_0)$$

Probability of restart (t=0)

Prob of restart *in each layer / network*

Column-normalized adjacency mat defined by blocks:

Adj. for N_1	Bipartite int.
Bipartite int.	Adj. for N_2

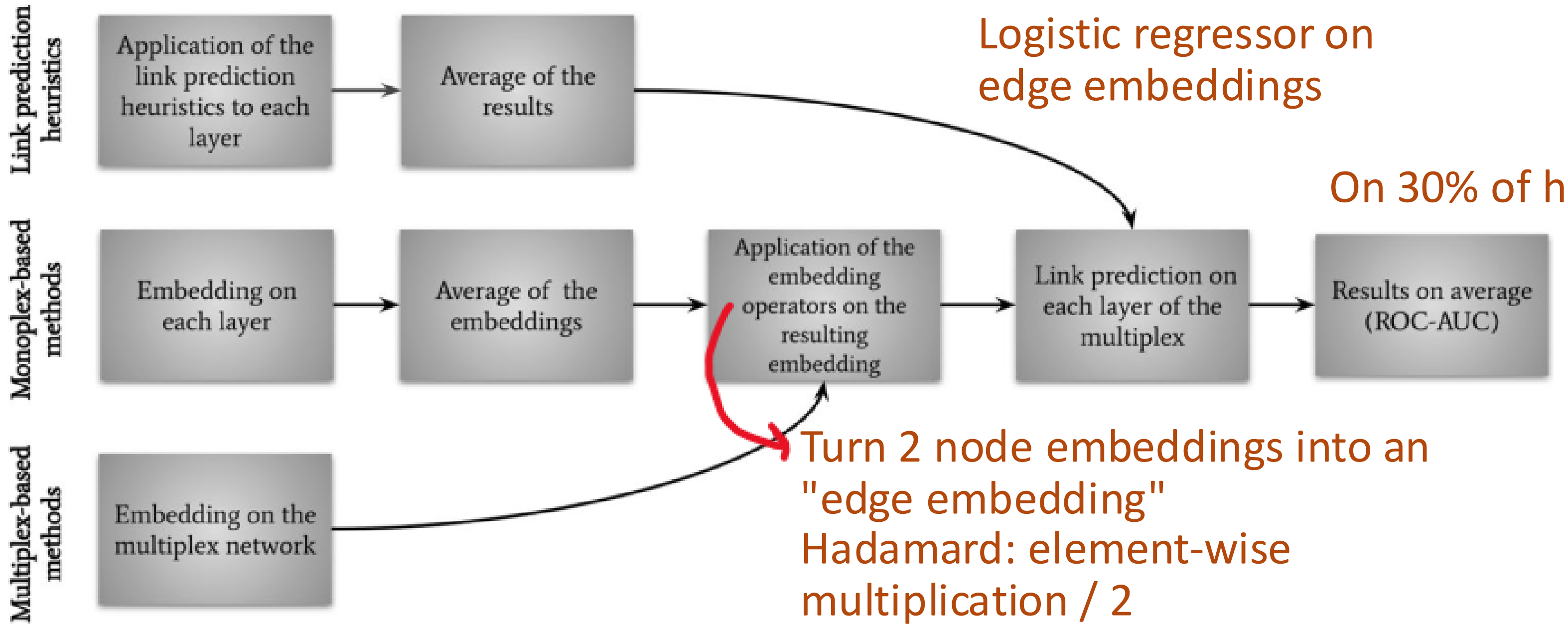
Apply RWR-MH n=#nodes(N_1)+#nodes(N_2) times

➡ n x n matrix for simG(.,.)

δ=0.5 r=0.7 $\underline{\boldsymbol{\tau}}$ uniform

1. **Learn node similarities with a random walk specific to multiplex(-het) networks**
2. Learn embeddings with VERSE

Experimental results show applications to link prediction & network reconstruction (showing LP).



Logistic regressor on edge embeddings

On 30% of held-out edges

Turn 2 node embeddings into an "edge embedding"
Hadamard: element-wise multiplication / 2

From the MultiVERSE paper.

Experimental results show applications to link prediction & network reconstruction (showing LP).

Multiplex networks

Method	CKM	LAZEGA	C.ELE	ARXIV	DIS	HOMO	MOL
node2vec-av	0.7908	0.6372	0.8552	0.9775	0.9093	0.8638	0.8753
deepwalk-av	0.7467	0.6301	0.8574	0.9776	0.9107	0.8638	0.8763
LINE-av	0.5073	0.4986	0.5447	0.8525	0.9013	0.8852	0.8918
Ohmnet	0.7465	0.7981	0.833	0.9605	0.9333	0.9055	0.8613
MNE	0.5756	0.6356	0.794	0.9439	0.9099	0.8313	0.8736
Multi-node2vec	0.8182	0.7884	0.8375	0.9581	0.8528	0.8592	0.8835
MultiVERSE	0.8177	***0.8269***	***0.8866***	***0.9937***	***0.9401***	***0.917***	***0.9259***

From the MultiVERSE paper.

Monoplex-net methods

Multiplex-net methods

Experimental results show applications to the recovery/discovery of gene-disease associations for Progeria.

MultiVERSE on multiplex-heterogeneous gene-disease and drug-target networks

Operators	Gene-disease bipartite	Drug-target bipartite
Hadamard	**0.95**	**0.9701**
Weighted-L1	0.7962	0.8057
Weighted-L2	0.7951	0.8055
Average	**0.9603**	**0.9703**
Cosine	0.7765	0.8338

From the MultiVERSE paper.

Perspectives

1. Comments on the paper
2. Why is it interesting for BioComp?

My comments on the paper

Strengths:
- Easy to understand and good performance

(did a lot of optimization for the VERSE framework to adapt it to large networks)
- Comprehensive experimental results
- Open-source

Weaknesses:
- Choice of parameters for the random walk
- Scalability: $O(N^2)$ where N = #nodes
- Combination from prior works (VERSE and RWR-MH)

Your comments?

Why is it interesting for BioComp?

Bridging the gap between imaging and tabular data for deep learning

(I think)

Should you care about Tabular AI?

SOTA #7: TabPFN, CARTE, and few-shot data science

MARIE BRAYER
MAR 06, 2025

7 4 Share

https://ontheflyinvesting.substack.com/p/should-you-care-about-tabular-ai

Using networks might be a good way to integrate prior knowledge information to imaging data analysis

➡ Embeddings are a way to capture information from networks