Network multiplication and derived networks

Vladimir Batagelj

IMFM Ljubljana, IAM UP Koper, and NRU HSE Moscow

Net 2021

on Zoom, October 18-20, 2021

Outline

Vladimir Batagelj: vladimir.batagelj@fmf.uni-lj.si
Current version of slides (October 19, 2021 at 06 : 43): slides PDF
https://github.com/bavla/NormNet/tree/main/docs

Networks / Formally

A *network* $\mathcal{N} = (\mathcal{V}, \mathcal{L}, \mathcal{P}, \mathcal{W})$ consists of:

- a *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{L})$, where $\mathcal{V}$ is the set of nodes, $\mathcal{A}$ is the set of arcs, $\mathcal{E}$ is the set of edges, and $\mathcal{L} = \mathcal{E} \cup \mathcal{A}$ is the set of links.
  $n = |\mathcal{V}|$, $m = |\mathcal{L}|$
- $\mathcal{P}$ *node value functions* / properties: $p : \mathcal{V} \to A$
- $\mathcal{W}$ *link value functions* / weights: $w : \mathcal{L} \to B$

Two-mode networks

In a *two-mode* network $\mathcal{N} = ((\mathcal{U}, \mathcal{V}), \mathcal{L}, \mathcal{P}, \mathcal{W})$ the set of nodes consists of two disjoint sets of nodes $\mathcal{U}$ and $\mathcal{V}$, and all the links from $\mathcal{L}$ have one endnode in $\mathcal{U}$ and the other node in $\mathcal{V}$. Often also a *weight* $w : \mathcal{L} \to \mathbb{R} \in \mathcal{W}$ is given; if not, we assume $w(u, v) = 1$ for all $(u, v) \in \mathcal{L}$ – a *binary* network.

A two-mode network can also be described by a rectangular matrix $\mathbf{A} = [a_{uv}]_{\mathcal{U} \times \mathcal{V}}$.

$$a_{uv} = \begin{cases} w(u, v) & (u, v) \in \mathcal{L} \\ 0 & \text{otherwise} \end{cases}$$

Multirelational and linked networks

In a *linked* or *multimodal* network

$$\mathcal{N} = ((\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_j), (\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_k), \mathcal{P}, \mathcal{W})$$

the set of nodes $\mathcal{V}$ is partitioned into subsets (*modes*) $\mathcal{V}_i$, and the set of links $\mathcal{L}$ is partitioned into subsets (*relations*) $\mathcal{L}_i$, $\mathcal{L}_s \subseteq \mathcal{V}_p \times \mathcal{V}_q$. Properties and weights are usually partial functions.

A linked network can be described also as a *collection* of one/two-mode networks $(\mathcal{N}_k)_{k \in K}$.

Multiplication of networks

Let **UV** (on sets U and V) and **VZ** (on sets V and Z) be matrices of the corresponding two-mode networks $\mathcal{N}_{UV} = ((U, V), L_{UV}, w_{UV})$ and $\mathcal{N}_{VZ} = ((V, Z), L_{VZ}, w_{VZ})$; L – set of links, w – weight on links. Their product $\mathbf{UZ} = \mathbf{UV} \cdot \mathbf{VZ}$

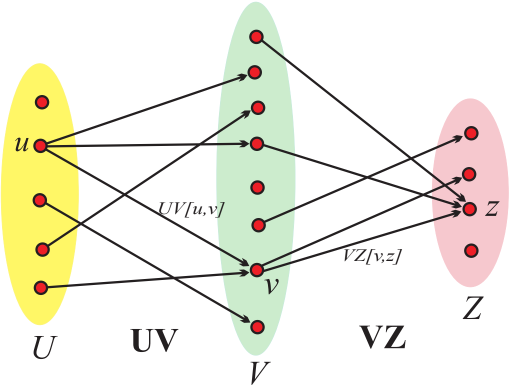
$$UZ[u, z] = \sum_{v \in V} UV[u, v] \cdot VZ[v, z]$$

determines the corresponding *product network* $\mathcal{N}_{UZ} = ((U, Z), L_{UZ}, w_{UZ}) = \mathcal{N}_{UV} \cdot \mathcal{N}_{VZ}$.

The definition can be extended to semirings [Cerinšek and Batagelj(2017)].

What is the complexity of computing the product of large sparse networks [Batagelj and Cerinšek(2013), Batagelj et al.(2014)] ?

Multiplication of networks meaning



If networks $\mathcal{N}_{UV}$ and $\mathcal{N}_{VZ}$ are binary the value of $UZ[u,z] = |N_{UV}(u) \cap N_{ZV}(z)|$ (N – set of neigbors) counts the number of ways we can go from $u \in U$ to $z \in Z$ passing through V.

Outer product decomposition

For vectors $x = [x_1, x_2, \ldots, x_n]$ and $y = [y_1, y_2, \ldots, y_m]$ their *outer product* $x \circ y$ is defined as a matrix $x \circ y = [x_i \cdot y_j]_{n \times m}$. The product $\mathbf{UZ}$ can be expressed as the *outer product decomposition* [Batagelj(2020)]

$$\mathbf{UZ} = \mathbf{UV} \cdot \mathbf{VZ} = \sum_v \mathbf{H}_v \quad \text{where} \quad \mathbf{H}_v = \mathbf{UV}[\cdot, v] \circ \mathbf{VZ}[v, \cdot]$$

For binary matrices we have $\mathbf{H}_v = K_{N_{UV}(v), N_{VZ}(v)}$ – the product is a sum of complete bipartite subgraphs.

Let $T(\mathbf{UV}) = \sum_u \sum_v UV[u, v]$ denote the *total weight* of the network $\mathcal{N}_{UV}$. Then

$$T(\mathbf{UZ}) = \sum_v T(\mathbf{H}_v) \quad \text{where} \quad T(\mathbf{H}_v) = \mathrm{wid}_{UV}(v) \cdot \mathrm{wod}_{VZ}(v),$$

$$\mathrm{wid}_{UV}(v) = \sum_u UV[u, v] \quad \text{and} \quad \mathrm{wod}_{VZ}(v) = \sum_z VZ[v, z].$$

Fast multiplication of sparse networks

Most of large networks are sparse (Dunbar's number). We can multiply sparse networks faster considering only nonzero elements [Batagelj et al.(2014)]

for v **in** V **do**
  for $(u, z) \in N_{VU}(v) \times N_{VZ}(v)$ **do**
    if $\exists UZ[u, z]$ **then** $UZ[u, z] := UZ[u, z] + UV[u, v] \cdot VZ[v, z]$
    else new $UZ[u, z] := UV[u, v] \cdot VZ[v, z]$

$M(code)$ = the number of multiplications in the execution of *code*.
$M(\mathbf{H}_v) = \mathrm{id}_{UV}(v) \cdot \mathrm{od}_{VZ}(v)$, $\quad A = \max_v \mathrm{id}_{UV}(v) \cdot \mathrm{od}_{VZ}(v)$

$$A \leq M(\mathbf{UZ}) \leq |V|A$$

For $v \in V$ such that $\mathrm{id}_{UV}(v), \mathrm{od}_{VZ}(v) \sim O(n)$ we have $A \sim O(n^2)$.
$M(\mathbf{UZ})$ has at least quadratic complexity.

Criteria for fast multiplication

For a sequence (in decreasing order) $\mathbf{d} = (d_i)_{i \in I}$

$$WP(\mathbf{d}) = \arg\min_i \{i \in I : d_i < i\}$$

is called its *Welsh-Powel number* [Welsh and Powell(1967)].
$wp_U = WP(\mathrm{id}_{UV})$ and $wp_Z = WP(\mathrm{od}_{VZ})$;
$WP_U = \{v \in V : \mathrm{id}(v) \geq wp_U\}$ and $WP_Z = \{v \in V : \mathrm{od}(v) \geq wp_Z\}$
$\Delta_U = \max_{v \in V} \mathrm{id}(v)$ and $\Delta_Z = \max_{v \in V} \mathrm{od}(v)$
The set V is partitioned into WP_U, WP_Z and $V \setminus (WP_U \cup WP_Z)$

$$M(\mathbf{UZ}) \leq wp_U \cdot wp_Z \cdot (\Delta_U + \Delta_Z + |V| - (wp_U + wp_Z))$$

$\Delta_U \geq wp_U$ and $\Delta_Z \geq wp_Z \Rightarrow M(\mathbf{UZ}) \leq wp_U \cdot wp_Z \cdot O(|V|)$

Theorem. If $|U|, |V|, |Z| \sim O(n)$, $wp_U \cdot wp_Z \sim O(1)$ and $WP_U \cap WP_Z = \emptyset$ then $M(\mathbf{UZ}) \sim O(n)$.

Derived networks

The product provides a new network linking set U to set Z. In special cases (*projections*), $\mathbf{UU} = \mathbf{UV} \cdot \mathbf{UV}^T$ and $\mathbf{VV} = \mathbf{UV}^T \cdot \mathbf{UV}$, it transforms a two-mode network to an ordinary (one-mode) network. This turns out to be very useful in the analysis of collections of networks.

Example:
U = set of authors, V = set of papers, $u\mathbf{UV}v \equiv u$ coauthored v
Z = set of keywords, $v\mathbf{VZ}z \equiv v$ is described by z
$UZ[u, z] = \#$ of papers authored by u described by z.
$UU[u, t] = \#$ of papers coauthored by u and t.

Derived networks

To get the right answers to some questions we have often to normalize networks used in products [Batagelj(2020)]. The contributions of intermediate nodes v to the product depend on sizes and values in $\mathbf{H}_v$.

Networks obtained from basic networks from a collection using multiplications or normalizations are called *derived* networks.

Assume that we would like that each intermediate node $v \in V$ contributes the same to the product total, $T(\mathbf{H}'_v) = 1$, for normalized networks $\mathbf{UZ}' = n(\mathbf{UV}) \cdot n(\mathbf{VZ})$. We have

$$T(\mathbf{H}'_v) = \mathrm{wid}_{n(\mathbf{UV})}(v) \cdot \mathrm{wod}_{n(\mathbf{VZ})}(v) = 1$$

This can be achieved if we set $\mathrm{wid}_{n(\mathbf{UV})}(v) = 1$ and $\mathrm{wod}_{n(\mathbf{VZ})}(v) = 1$. Assume that $n(\mathbf{UV})[u,v] = UV[u,v]/S, S > 0$. We get

$$\mathrm{wid}_{n(\mathbf{UV})}(v) = \sum_u n(\mathbf{UV})[u,v] = \sum_u UV[u,v]/S = \mathrm{wid}_{UV}(v)/S = 1$$

Stochastic normalization
Fractional approach

For $\mathrm{wid}_{UV}(v) > 0$ and $\mathrm{wod}_{VZ}(v) > 0$ we finally have

$$n(UV)[u,v] = \frac{UV[u,v]}{\mathrm{wid}_{UV}(v)} \quad \text{and} \quad n(VZ)[v,z] = \frac{VZ[v,z]}{\mathrm{wod}\,UV(v)}$$

Note: in general we can have nodes with in/out-degree 0.

$$V^+(f) = \{v \in V : f(v) > 0\}$$

$$T(n(\mathbf{UV})) = \sum_v \sum_u n(\mathbf{UV})[u,v] = \sum_{v:\mathrm{id}(v)>0} 1 = V^+(\mathrm{id}_{UV}(v))$$

and

$$T(\mathbf{UZ}') = \sum_v T(\mathbf{H}'_v) = |V^+(\mathrm{wid}_{UV}(v) \cdot \mathrm{wod}_{VZ}(v))|$$

Each active node in V has value 1 which is distributed over links from U to Z.

Normalized projection

In a special case $\mathbf{VZ} = \mathbf{UV}^T = \mathbf{VU}$, $\mathbf{VV} = \mathbf{VU} \cdot \mathbf{UV}$ we have

$$\mathrm{wid}_{UV}(v) = \mathrm{wod}_{UV^T}(v) = \mathrm{wod}_{VZ}(v) = \mathrm{wod}_{VU}(v)$$

and

$$n(VU)[v,u] = \frac{VU[v,u]}{\mathrm{wod}_{VU}(v)} \quad \text{and} \quad n(\mathbf{UV}) = n(\mathbf{VU})^T$$

$$\mathbf{VV}' = n(\mathbf{VU}) \cdot n(\mathbf{UV}) = n(\mathbf{VU}) \cdot n(\mathbf{VU})^T$$

Example:
V = set of papers, U = set of authors, $u\mathbf{VU}v \equiv v$ is an author of u
$2 \cdot \mathbf{VV}'[v,x]$ = fractional contribution of collaboration of authors v and x to the bibliography
Newman's normalization [Newman(2004)].

Linking through a network

Assume that we have an additional ordinary network $\mathbf{S}$ on V. We say that

$$\mathbf{UZ}_S = \mathbf{UV} \cdot \mathbf{S} \cdot \mathbf{VZ}$$

links U to Z *through* $\mathbf{S}$.

In the bibliographic example we can consider $\mathbf{S}$ = citation network. Then $\mathbf{UZ}_S[u, z]$ = number of citations of the author u to works described by keyword z.

As a fractional version we consider

$$\mathbf{UZ}'_S = n(\mathbf{UV}) \cdot \mathbf{S} \cdot n(\mathbf{VZ})$$

Let's look at its total (all nodes in V active)

$$T(\mathbf{UZ}'_S) = \sum_{v,y} S[v, y] \cdot \sum_u n(\mathbf{UV})[u, v] \cdot \sum_z n(\mathbf{VZ})[y, z] = T(\mathbf{S})$$

Linking through a network

In the network $\mathbf{UZ}'_S$ the total value of the network $\mathbf{S}$ is redistributed on links from U to Z.

Replacing $\mathbf{S}$ with $n(\mathbf{S})$ we get

$$T(\mathbf{UZ}'_{n(S)}) = T(n(\mathbf{S})) = V^+(\mathrm{od}_S(v))$$

In the network $\mathbf{UZ}'_{n(S)}$ each active node $v \in V$ has value 1 which is redistributed on links from U to Z.

Alternative normalized projections

Let $b(\mathbf{UV})$ be a *binarized* version of a network $\mathbf{UV}$. $b(\mathbf{UV})[u,v] = \delta((u,v) \in \mathcal{L}_{UV})$. For a network $\mathbf{UV}$ we define the *left fractional contribution* $l(\mathbf{UV}) = n(\mathbf{UV})^T \cdot b(\mathbf{UV})$ and the *right fractional contribution* $r(\mathbf{UV}) = b(\mathbf{UV})^T \cdot n(\mathbf{UV})$. We have

$$r(\mathbf{UV})^T = (b(\mathbf{UV})^T \cdot n(\mathbf{UV}))^T = n(\mathbf{UV})^T \cdot b(\mathbf{UV}) = l(\mathbf{UV})$$

and

$$l(UV)[v,y] = \frac{1}{\mathrm{wod}(v)} \sum_{u \in U} VU[v,u] \cdot b(VU)[y,u] \leq 1$$

In general $l(UV)[v,y] \neq r(UV)[v,y]$. A symmetric measure/ projection $VV_X[u,t]$ can be constructed as some $mean_X$ of these two quantities

$$\begin{aligned} VV_X[v,y] &= mean_X(l(UV)[v,y], r(UV)[v,y]) \\ &= mean_X(l(UV)[v,y], l(UV)[y,v]) \end{aligned}$$

Alternative normalized projections

$$VV_A[v,y] = \frac{1}{2}(I(UV)[v,y] + I(UV)[y,v]) \text{ – arithmetic mean}$$
$$VV_m[v,y] = \min(I(UV)[v,y], I(UV)[y,v]) \text{ – minimum}$$
$$VV_M[v,y] = \max(I(UV)[v,y], I(UV)[y,v]) \text{ – maximum}$$
$$VV_G[v,y] = \sqrt{I(UV)[v,y] \cdot I(UV)[y,v]} \text{ – geometric mean, Salton}$$
$$VV_H[v,y] = 2(I(UV)[v,y]^{-1} + I(UV)[y,v]^{-1})^{-1} \text{ – harmonic mean, Dice}$$
$$VV_J[v,y] = (I(UV)[v,y]^{-1} + I(UV)[y,v]^{-1} - 1)^{-1} \text{ – Jaccard}$$

$$VV_X[v,y] = VV_X[y,v]$$

$$VV_X[v,y] \in [0,1]$$

$$VV_J[v,y] \leq VV_m[v,y] \leq VV_H[v,y] \leq VV_G[v,y] \leq VV_A[v,y] \leq VV_M[v,y]$$

Alternative normalized projections of binary networks

$$VV_A[v,y] = \frac{|VU(v) \cap VU(y)|}{2}\left(\frac{1}{|VU(v)|} + \frac{1}{|VU(y)|}\right)$$

$$VV_m[v,y] = \frac{|VU(v) \cap VU(y)|}{\max(|VU(v)|, |VU(y)|)}$$

$$VV_M[v,y] = \frac{|VU(v) \cap VU(y)|}{\min(|VU(v)|, |VU(y)|)}$$

$$VV_G[v,y] = \frac{|VU(v) \cap VU(y)|}{\sqrt{|VU(v)| \cdot |VU(y)|}}$$

$$VV_H[v,y] = \frac{2|VU(v) \cap VU(y)|}{|VU(v)| + |VU(y)|}$$

$$VV_J[v,y] = \frac{|VU(v) \cap VU(y)|}{|VU(v) \cup VU(y)|}$$

$$|VU(v) \cup VU(y)| = |VU(v)| + |VU(y)| - |VU(v) \cap VU(y)|$$

$|VU(v) \cap VU(y)| = VV[u,t]$ and $|VU(v)| = VV[v,v]$

Additional references I

- Batagelj, V., Cerinšek, M. (2013). On bibliographic networks. Scientometrics 96 (3), 845-864.
- Batagelj, V., Doreian P., V., Ferligoj, A., Kejžar N. (2014). Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution. Wiley.
- Batagelj, V. (2020). On fractional approach to analysis of linked networks. Scientometrics 123 (2) 621–633.
- Batagelj, V. (2021). Analysis of the Southern women network using fractional approach. Social Networks 68, 229–236.
- Cerinšek, M., Batagelj, V. (2017). Semirings and Matrix Analysis of Networks. in Encyclopedia of Social Network Analysis and Mining. Reda Alhajj, Jon Rokne (Eds.), Springer.

Additional references II

Derived Networks

V. Batagelj

- Newman, M.E.J. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences of the United States of America 101(Suppl1), 5200–5205.
- Welsh, D. J. A., Powell, M. B. (1967). An upper bound for the chromatic number of a graph and its application to timetabling problems. The Computer Journal, 10 (1): 85–86