Indirect Influences

Rafael Díaz

Abstract

We introduce PWP a method for counting indirect influences and compare it with a couple of well-known methods. We provide combinatorial as well as probabilistic interpretation for the PWP method.

1 Introduction

The goal of this note is to compare three alternative approaches to study indirect influences. Two of these approaches, namely the MICMAC and PageRank methods, are already well-tested and each possesses a host of applications. We believe that the third method PWP, while still in its infancy, may also be useful in a variety of contexts. We consider a discrete scenario where the variables involved are identified with set \([n] = \{1, 2, ..., n\}\). We assume that direct influences can be encoded into a \(\mathbb{R}\)-weighted directed graph without multiple edges with \([n]\) as its set of vertices. Thus, a direct influence from the variable \(j\) on the variable \(i\) corresponds with an edge from vertex \(j\) to vertex \(i\). As usual we encode the graph into the matrix \(D\) of direct influences such that \(D_{i,j}\) is the weight of the edge going from \(j\) to \(i\), if there is one, and \(D_{i,j} = 0\) otherwise. Our problem consists in evaluating the indirect influences between vertices, i.e. finding the matrix \(T\) coding indirect influences. In this work we make a comparative analysis of three different definitions for \(T\).

2 MICMAC

The MICMAC method introduced by Godet \cite{Godet} works as follows. Let \(D\) be the matrix associated with the graph of direct influences. The graph of indirect influences is represented by the matrix \(T = D^k\), where \(k\) is a fixed small natural number, say 4 or 5. One defines the vector of dependencies \(d = (d_1, ..., d_n)\) and the vector of influences \(f = (f_1, ..., f_n)\), where the coefficients \(d_j\) and \(f_j\) are given by:

\[
d_j = \sum_{i=1}^{n} T_{ji} \quad \text{and} \quad f_j = \sum_{i=1}^{n} T_{ij}. \tag{1}
\]

Thus \((d_j, f_j)\) is equal to the \((in, out)\)-degree of the vertex \(j\) in the graph of indirect influences. The numbers \(d_j\) and \(f_j\) encode valuable information, for example, the most influential vertex \(j\) is the one with the highest coefficient \(f_j\).


3 PageRank

The celebrated method PageRank [5] for studying indirect dependencies has already revolutionized the world of internet search engines through the remarkable success of Google. The mathematics encoding the basic structures behind PageRank are however surprisingly simple. With hindsight we observe that PageRank differs from MICMAC in three main ideas: 1) Normalization of influences, i.e. use of column stochastic matrices. 2) Use of complete graphs, i.e. graphs whose associated matrix have no vanishing entries. 3) Taking infinite potencies of matrices.

Let $D$ be the $n \times n$ matrix of direct influences. It is assumed that the entries of $D$ are non-negative real numbers, and that the sum of the entries of each column of $D$ is either 0 or 1. With PageRank dependencies and influences are determined by the pairs $(d_i, f_i)$ given by (1) where now the matrix $T$ of indirect influences is computed as follows:

$$T = \lim_{k \to \infty} \left[ p\overrightarrow{D} + (1 - p)E_n \right]^k$$

where: 1) $0 < p < 1$ is a chosen number close to 1, say $p = 0.86$. 2) $\overrightarrow{D}$ is obtained from $D$ by replacing the entries of each zero column of $D$ by $\frac{1}{n}$. 3) $E_n$ is the matrix with entries equal to $\frac{1}{n}$.

For the web application $D$ is constructed as follows. Consider the web graph whose set of vertices $[n]$ represents the set of all web pages in the world wide web. There is an edge from page $j$ to page $i$ in the web graph if and only if a link to page $i$ appears in the content of page $j$. The matrix $D$ of direct influences is such that $D_{i,j} = \frac{1}{\text{out}(j)}$ if there is an edge from $j$ to $i$, and $D_{i,j} = 0$ otherwise; $\text{out}(j)$ is the out-degree of the vertex $j$ in the web graph. Since $p\overrightarrow{D} + (1 - p)E_n$ is a column stochastic matrix, the vector of influences for $T$ is $(1, ..., 1)$; thus the relevant information is now in the vector of dependencies. The entries of the vector $\frac{1}{n}(d_1, ..., d_n)$ yields the PageRank number of each web page. The greater the PageRank number of a page the greater its importance. By construction $p\overrightarrow{D} + (1 - p)E_n$ is the transition matrix of a primitive irreducible Markov chain [4]. Therefore the vector of indirect dependencies $d = (d_1, ..., d_n)$ is an eigenvector of $T$ with eigenvalue 1, that is $Td^t = d^t$. Moreover, the limit (2) can be computed in terms of $d$ as follows $T = (d^t \ d^t \ ... \ d^t)$.

4 PWP

We introduce the PWP method for counting indirect influences. PWP can be applied to any matrix of direct influences, and while MICMAC focuses on paths of length $k$, and PageRank focuses on infinite long paths, PWP takes into account paths of various lengths.

In a nutshell the PWP method can described as follows: given the matrix of direct influences \( D \), fixed a real number \( \lambda > 0 \) and let \( T \), the matrix of indirect influences, be given by

\[
T = T(D) = \frac{1}{e^\lambda} e^{\lambda D} \quad \text{where} \quad e^x = e^x - 1 = \sum_{k=1}^{\infty} \frac{x^k}{k!}.
\]

Notice that \( D^k = \frac{\partial}{\partial \lambda} (e^\lambda T) |_{\lambda=0} \), and thus, in principle, we can compute the MICMAC matrix of indirect influences from the PWP matrix of indirect influences. As an example, let \( L_2 \) be the graph with vertex set \([2]\) and one edge from vertex 1 to vertex 2. The only non-vanishing entry of \( T \) is \( T_{1,2} = \lambda e^\lambda - 1 \), making vertex 1 the most influential vertex. Also we get that vertex 2 does not exert any influence whatsoever over vertex 1. The indirect influence of vertex 1 over vertex 2 is given by

\[
\lambda e^\lambda - 1 = \sum_{k=0}^{\infty} B_k \frac{\lambda^k}{k!},
\]

where the coefficients \( B_k \) are the so-called Bernoulli numbers. See [2] for an explicit definition and a combinatorial interpretation for the Bernoulli numbers. Notice that \( \frac{\lambda}{e^\lambda - 1} \) is less than 1, it approaches 1 as \( \lambda \) approaches 0, and it approaches 0 as \( \lambda \) goes to infinity. Therefore the indirect influence that 1 exerts over 2 is less than the original direct influence, it approaches its original value when \( \lambda \) approaches 0, and it is negligible if \( \lambda \) is a large number. The direct influence that 1 exerts over 2 becomes less relevant as \( \lambda \) increases, since with PWP only paths of length close to \( \lambda \) are really taken into account.

Our first result states some of the basic properties of the map \( D \rightarrow T(D) \) and provides a probabilistic interpretation for the PWP method.

**Theorem 1.**
1. \( T(D^t) = T(D)^t \), and if \( D = QCQ^{-1} \) then \( T(D) = QT(C)Q^{-1} \).
2. Let \( D_1 \oplus D_2 \in M_m(\mathbb{R}) \oplus M_n(\mathbb{R}) \subseteq M_{m+n}(\mathbb{R}) \), then \( T(D_1 \oplus D_2) = T(D_1) \oplus T(D_2) \).
3. If \( D \) is a column stochastic matrix, then \( T(D) \) is a column stochastic matrix.
4. If \( [D_1, D_2] = 0 \), then we have that

\[
T(D_1 + D_2) = e^\lambda T(D_1)T(D_2) + T(D_1) + T(D_2).
\]

5. If \( D \in M_n(\mathbb{R}) \) then \( T(D) \) is the expected matrix of \( \widehat{D} \), where \( \widehat{D} : \mathbb{N}_+ \rightarrow M_n(\mathbb{R}) \) is the random matrix given by \( \widehat{D}(k) = D^k \), and \( \mathbb{N}_+ \) is the probability space with probability function given by \( p(k) = \frac{\lambda^k}{e^\lambda - 1} \).

Let \( P \) be the Poisson probability on \( \mathbb{N} \) given by \( P(k) = e^{-\lambda} \frac{\lambda^k}{k!} \).

**Proposition 2.**
1. For \( k \geq 1 \) we have that \( p(k) = \frac{P(k)e^\lambda}{e^\lambda - 1} \). If \( X \) is a random variable with distribution \( p \), then \( EX = \frac{\lambda e^\lambda}{e^\lambda - 1} \), \( EX^2 = \frac{(\lambda^2 + \lambda)e^\lambda}{e^\lambda - 1} \), \( V X = \frac{\lambda e^\lambda - (\lambda^2 + \lambda)e^\lambda}{(e^\lambda - 1)^2} \).
Therefore, by the Chebyshev’s inequality, we have that

\[ p \left( \left| X - \frac{\lambda e^\lambda}{e^\lambda - 1} \right| \geq c \right) \leq \frac{\lambda e^{2\lambda} - (\lambda^2 + \lambda)e^\lambda}{e^\lambda (e^\lambda - 1)^2}. \]

In order to provide a combinatorial interpretation for the PWP method we need a few definitions \([\Pi]\). Let \(\mathbb{R}\)-Set be the category of \(\mathbb{R}\)-weighted sets, i.e. the category whose objects are pairs \((x, \omega)\) where \(x\) is a set and \(\omega : x \to \mathbb{R}\) is a map. A morphisms in \(\mathbb{R}\)-Set from \((x_1, \omega_1)\) to \((x_2, \omega_2)\) is a map \(\alpha : x_1 \to x_2\) such that \(\omega_1 = \omega_2 \circ \alpha\). We also consider \(\mathbb{R}\)-set the full subcategory of \(\mathbb{R}\)-Set whose objects are pairs \((x, \omega)\) \(\in \mathbb{R}\)-Set where \(x\) is a finite set. \(\mathbb{R}\)-set is a distributive category provided with a valuation map \(|_\cdot| : \mathbb{R}\text{-set} \to \mathbb{R}\) given by

\[ |x, \omega| = \sum_{i \in x} \omega(i). \]

If \(e\) is an edge of a directed graph we denote by \(s(e)\) and \(t(e)\) the starting point and the endpoint of \(e\), respectively. A path \(\gamma\) of length \(k\) from a vertex \(j\) to a vertex \(i\) in a graph is a sequence of edges \(\gamma = (\gamma_1, \ldots, \gamma_k)\) such that \(s(\gamma_1) = j, t(\gamma_i) = s(\gamma_{i+1})\) for \(1 \leq i < k\) and \(t(\gamma_k) = i\). We denote by \(P(i, j)\) the set of all paths from \(j\) to \(i\), and by \(P_k(i, j)\) the set of all paths of length \(k\) from \(j\) to \(i\).

We assume that the graph of direct influences has associated matrix \(D\). In our applications we use the \(\mathbb{R}\)-weighted sets \((P_k(i, j), \omega)\), \((P(i, j), \rho)\) and \((P(i, j), \omega_\lambda)\) where \(\omega\), \(\rho\), and \(\omega_\lambda\) are given, respectively, on a path \(\gamma = (\gamma_1, \ldots, \gamma_k)\) in the graph of direct influences by

\[
\omega(\gamma) = \prod_{i=1}^{k} D(t(\gamma_i), s(\gamma_i)), \quad \rho(\gamma) = \prod_{i=1}^{k} (pD + (1 - p)E_n)_{t(\gamma_i), s(\gamma_i)} , \quad \omega_\lambda(\gamma) = \left( \prod_{i=1}^{k} D(t(\gamma_i), s(\gamma_i)) \right) \frac{\lambda^k}{e^\lambda k!}.
\]

The following result provides combinatorial interpretation for the MICMAC, PageRank, and the PWP methods.

**Theorem 3.** 1. If \(T\) is the MICMAC matrix of indirect influences, then \(T_{ij} = |P_k(i, j), \omega|\).
2. If \(T\) is the PageRank matrix of indirect influences, then \(T_{ij} = \lim_{k \to \infty} |P_k(i, j), \rho|\).
3. If \(T\) is the PWP matrix of indirect influences, then \(T_{ij} = |P(i, j), \omega_\lambda| = \sum_{k=1}^{\infty} |P_k(i, j), \omega| \frac{\lambda^k}{e^\lambda k!}\).

**5 Examples**

**Example 4.** Let \(L_n\) be the graph with vertex set \([n]\) and with an edge from \(i\) to \(i + 1\) for \(i < n\). According to MICMAC, for \(k < n\), vertex \(i\) will only influence vertex \(i + k\). Thus the vertices \(\{1, \ldots, n - k\}\) are the most influential ones each with influence 1. For \(k \geq n\) the matrix of indirect influences vanishes. MICMAC for \(n = 3, k = 4\) predicts vanishing influences and dependencies.
PageRank for \( n = 3 \) gives the dependency vector \( (0.17, 0.34, 0.47) \) making vertex 3 the most dependent one. PWP for \( n = 3 \) and \( \lambda = 1 \) gives the vectors \( f = (1.5, 1, 0) \) and \( d = (0, 1, 1.5) \). Vertex 3 is the most dependent and vertex 1 has vanishing dependency. Similarly, MICMAC for \( n = 4, k = 4 \) predicts vanishing influences, PageRank dependencies are \( (0.12, 0.21, 0.30, 0.37) \) and PWP with \( \lambda = 1 \) yields the vectors \( f = (\frac{1}{3}, \frac{1}{3}, 1, 0) \) and \( d = (0, 1, \frac{1}{2}, \frac{1}{2}) \). PageRank lowers the dependency of vertex 1 to 0.17, 0.12, 0.08, 0.059, 0.046 as \( n \) moves from 3 to 7. Thus for large \( n \) we expect PageRank to produce an almost vanishing dependency for vertex 1, thus approaching the PWP’s result. PWP for arbitrary \( \lambda > 0 \) yields the matrix \( T \) given by \( T_{j+s,j} = \frac{\lambda^s}{e^\lambda s!} \) for \( 1 \leq j < n, 1 \leq s \leq n-j \), and \( T_{j+s,j} = 0 \) otherwise. Notice that for \( s \leq n-j-1 \) we have that

\[
\frac{T_{j+s+1,j}}{T_{j+s,j}} = \frac{\lambda}{s+1},
\]

thus \( T_{j+s+1,j} > T_{j+s,j} \) if \( s < \lambda - 1 \) and \( T_{j+s+1,j} < T_{j+s,j} \) if \( s > \lambda - 1 \). Therefore \( T_{j+s,j} \) achieves its maximum, for fixed \( j \), at \( s = \lfloor \lambda \rfloor \) if \( \lambda \geq 1 \) and at \( s = 1 \) otherwise.

**Example 5.** Let \( C_n \) be the \( n \)-cycle graph on \([n]\), i.e. the graph with an edge from \( i \) to \( i+1 \) for \( i < n \) and an edge from \( n \) to 1. In this case the three methods yield the same vector of dependencies, but they do that for quite different reasons. MICMAC makes the influence and dependence of each vertex equal to 1. For fixed \( k \) vertex \( i \) will only influence vertex \((i+k)\) mod \( n \). PageRank vector of dependencies is also \((1, ..., 1)\), but in this case it means that each vertex is equally dependent on every other vertex. With PWP the matrix \( T \) of indirect influences is given, for \( j, s \in [n] \) and \( j + s \) taken mod \( n \), by

\[
T_{j+s,j} = \frac{1}{e^\lambda} \sum_{k=0}^\infty \frac{\lambda^{nk+s}}{(nk+s)!},
\]

so for \( \lambda = 1 \) we have \( T_{j+s,j} = \frac{1}{e^1} \sum_{k=0}^\infty \frac{1}{(nk+s)!} \)

and therefore we get that \( T_{j+1,j} > T_{j+2,j} > ... > T_{j+n,j} = T_{j,j} \), i.e. the total influence and dependence of vertex \( j \) is 1, and it influences all other vertices, but it has a higher influence over the vertices closer to it.

**Example 6.** According to properties 1 and 2 of Theorem 1, choosing and appropriated basis, one can always reduced the computation of \( T = T(D) \) to the case where \( D \) is a matrix in Jordan canonical form. Thus, we compute \( T \) for \( D \) a \( n \times n \) matrix such that \( D_{j,j} = a \in C, D_{j+1,j} = 1 \), and \( D_{i,j} = 0 \) otherwise. Vertex \( j \) exerts a non-vanishing indirect influence over the vertices \( j+s \) with \( 0 \leq s \leq n-j \). Notice that at vertex \( j \) a path can either stay at \( j \) or move to \( j+1 \), thus the MICMAC matrix is given by \(|P_k(j+s,j)| = (k)^{a^{k-s}}\), and therefore the PWP matrix of indirect influences is given by

\[
T_{j+s,s} = \frac{1}{e^\lambda} \sum_{k=s}^\infty \binom{k}{s} a^{k-s} \frac{\lambda^k}{k!} = \frac{e^{a\lambda} \lambda^s}{(e^\lambda - 1)s!}.
\]

Matrices in Jordan canonical form are not quite well-adapted for PageRank.
Example 7. Consider the star graph with vertex set \( \{0, 1, \ldots, n\} \) and edges from 0 to \( i \in [n] \) and vice versa. It is not hard to see that the PWP matrix of indirect influences \( T \) is given, for \( i, j \in [n] \), by

\[
T_{0,0} = \frac{e^{\lambda \sqrt{n}} + e^{-\lambda \sqrt{n}}}{2e^\lambda}, \quad T_{j,0} = T_{0,j} = \frac{1}{e^\lambda} \sum_{k=0}^{\infty} n^k \frac{\lambda^{2k+1}}{(2k+1)!}, \quad T_{i,j} = \frac{e^{\lambda \sqrt{n}} + e^{-\lambda \sqrt{n}}}{2n e^\lambda}.
\]

Thus the vertex with greater influence and dependence is the vertex 0. PageRank yields a similar result making 0 the most dependent vertex.

6 Final comments

We introduced the PWP method for counting indirect influences. Applications of PWP to real-world networks is currently underway. Our main goal in this note had been to propose an alternative route for counting indirect influences. It seems convenient to have a pool of options, as well as a comparative study of the various possibilities. We worked with a discrete scenario where influences are transmitted linearly. Lifting those restrictions will conduce to continuous non-linear models. This more general setting will be considered elsewhere.

References

[1] H. Blandón, R. Díaz, Rational combinatorics, Adv. in Appl. Math. 40 (2008) 107-126.

[2] R. Díaz, E. Pariguan, Super, Quantum and Non-Commutative Species, Afr. Diaspora J. Math. 8 (2009) 90-130.

[3] M. Godet, De l’Anticipation à l’Action, Dunod, Paris 1992.

[4] A. Langville, C. Meyer, Deeper Inside PageRank, Internet Mathematics 1 (2004) 335-400.

[5] S. Brin, R. Motwani, L. Page, T. Winograd, The PageRank citation ranking: Bringing order to the web, Technical Report, Stanford Digital Library Technologies Project, 1998.

ragadiaz@gmail.com
Facultad de Administración, Universidad del Rosario, Bogotá, Colombia