Differential Equivalence Yields Network Centrality

Stefano Tognazzi[1(✉)], Mirco Tribastone[1], Max Tschaikowski[1], and Andrea Vandin[2]

[1] IMT School for Advanced Studies, Lucca, Italy
stefano.tognazzi@imtlucca.it
[2] DTU Compute Copenhagen, Lyngby, Denmark

Abstract. One of the most distinctive features of collective adaptive systems (CAS) is the presence of many individuals which interact with each other and with the environment, giving rise to a system-level behaviour that cannot be analyzed by studying the single agents in isolation. The interaction structure among the individuals of CAS is often captured by networks where nodes denote individuals and edges interactions. Understanding the interplay between the network topology and the CAS dynamics calls for tools from network theory in order, for instance, to identify the most important nodes of a network. Centrality measures address this task by assigning an importance measure to each node, a possible example being the famous PageRank algorithm of Google. In this paper we investigate the relationship between centrality measures and model reduction techniques, such as lumpability of Markov chains, which seek to reduce a model into a smaller one that can be processed more efficiently, while preserving information of interest. In particular, we focus on the relation between network centrality and backward differential equivalence, a generalization of lumpability to general dynamical systems. We show that any two backward differential equivalent nodes enjoy identical centrality measures. By efficiently obtaining substantial reductions of real-world networks from biochemistry, social sciences and computer engineering, we demonstrate the applicability of the result.

Keywords: Networks · Centrality measures · Model reduction
Efficient algorithms

1 Introduction

More and more often we are facing systems consisting of a large number of entities, each with its own status, goals, and dynamics, which interact with each other (and with the environment) giving rise to an *emergent behaviour*, i.e. the system-level dynamics, that cannot be directly inferred by studying single individuals in isolation. Systems with such distinctive features are often referred to as collective adaptive systems (CAS). Given the importance of the interactions

T. Margaria and B. Steffen (Eds.): ISoLA 2018, LNCS 11246, pp. 186–201, 2018.
https://doi.org/10.1007/978-3-030-03424-5_13

in CAS, one possible approach to their analysis consists in conveniently representing the interaction structure in the form of a network, enabling a plethora of analysis techniques widely used in network theory [32].

Examples of networks arise in many real-world applications, with possible examples being collaboration, gene regulation, trust, internet and social networks. Identifying important components and structural properties of networks is crucial, and it has enjoyed substantial interest during the last decades, combining such diverse fields as graph theory, algebra and dynamical systems. Celebrated insights include, but are not limited to, small world phenomena and the concept of scale-free networks. Among many established notions, we focus on centrality measures [32] and exact role assignment [40].

Centrality measures are a common tool to identify nodes with a high impact in a network. Instead, exact role assignment (also known as regular equivalence), has been used since the late seventies [26,40] to formalize the idea "that nodes who occupy the same social position relate in the same ways with other nodes who are themselves in the same positions" [40]. Using the lingo of computer science, regular equivalence corresponds to the classic notion of bisimulation [2,25] and can be therefore computed efficiently by the partition refinement algorithm of Paige and Tarjan [33], whose running time is linear in the numbers of edges and logarithmic in the number of nodes. Surprisingly, it was not before 2003 that this fact has been observed [26], allowing one to avoid the usage of the cubic time algorithm CATREGE from 1993 [26].

The partition refinement algorithm of Paige and Tarjan was also key in the development of efficient model reduction algorithms for quantitative models [20,36]. The main idea behind model reduction is to relate the original dynamical model in a formal way to a reduced dynamical model such that a solution of the reduced model allows one to draw formal conclusions about the solution of the original model. In [16,38], the original algorithm [33] has been extended to efficiently compute the coarsest lumpable partition [7] of a Markov chain. In a similar vein of research, [2] presented an extension to probabilistic bisimulation in the style of Larsen and Skou [25]. The original algorithm [33] has been recently extended to efficiently compute the coarsest differential equivalence [10–12], a model reduction technique for systems of ordinary differential equations (ODEs) that are a natural modeling language for biochemical models [8,17], dependable systems [6,19] and analytical performance models [5,24]. In particular, [9,13,14] introduced a partition refinement algorithm that efficiently computes the largest differential equivalence.

The present work relates centrality measures, model reduction techniques and efficient algorithms. More specifically, by applying differential equivalence to the linear dynamical system that is induced by the adjacency matrix of a graph, we observe that any two differential equivalent nodes enjoy the same eigenvector, Katz and PageRank centrality. The relevance of the result is demonstrated by efficiently computing substantial reductions (that yields coarse quotient partitions with large blocks) of published real-world networks, including protein

interaction networks, scientific collaboration networks, routing networks, email networks and networks capturing YouTube.

Paper Outline. Section 2 reviews the background material, while Sect. 3 relates backward differential equivalence to centrality measures. Section 4, instead, features an extensive experimental evaluation on real-world networks. Section 5 concludes the paper.

2 Background

Notation. Let $\mathcal{V}$ be a finite index set and let $\mathbb{R}^{\mathcal{V}}$ denote the set of functions from $\mathcal{V}$ to $\mathbb{R}$. Elements of $\mathbb{R}^{\mathcal{V}}$ and $\mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ are called vectors and matrices, respectively. The set of variables is denoted by $\{x_i \mid i \in \mathcal{V}\}$. A partition of $\mathcal{V}$ is denoted by $\mathcal{H}$, its blocks by H. Partition $\mathcal{H}'$ refines partition $\mathcal{H}$ if for every $H' \in \mathcal{H}'$ there exists some $H \in \mathcal{H}$ such that $H' \subseteq H$.

2.1 Centrality Measures

Given a directed graph $G = (\mathcal{V}, E)$, the adjacency matrix $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ is obtained by setting $A_{i,j} = 1$ if $(i,j) \in E$ and $A_{i,j} = 0$ if $(i,j) \notin E$. Note that we can cover undirected graphs by ensuring that $(i,j) \in E$ if and only if $(j,i) \in E$. Our discussion on centrality measures follows standard literature, see for instance [32].

Eigenvector Centrality. For each node $i \in \mathcal{V}$, the corresponding eigenvector centrality x_i^* is defined as the average of eigenvector centralities of all nodes reachable from i, i.e., $x_i^* = \frac{1}{\lambda}\sum_{j\in\mathcal{V}} A_{i,j}x_j$, where $\frac{1}{\lambda}$ is some positive constant. More formally, one requires the following.

Definition 1. *The centrality measure of an adjacency matrix A is well-defined if there exists a unique $\lambda^* > 0$ and a unique non-negative $x^* \in \mathbb{R}^{\mathcal{V}}$ such that $Ax^* = \lambda^* x^*$ and $\|x^*\| = 1$.*

The centrality eigenvector is often computed by the power iteration method, introduced next.

Definition 2. *Fix a graph $G = (\mathcal{V}, E)$ and let $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ denote the underlying adjacency matrix. Then, given some nonzero vector $x^0 \in \mathbb{R}^{\mathcal{V}}$, the power iteration sequence $(x^k)_k$ is given by $x^{k+1} = Ax^k/\|x^k\|$ for all $k \geq 0$.*

The power iteration method is known to converge under the following common assumption.

(H) Assume that A has a unique largest eigenvalue λ^* that has a unique non-negative eigenvector x^* with $\|x^*\| = 1$.

Armed with **(H)**, the following can be proven if $x^0 = \mathbb{1}$, where $\mathbb{1}_i := 1$ for all $i \in \mathcal{V}$.

Theorem 1. *Under the assumption of* ***(H)****, the centrality eigenvector is well-defined and the power iteration sequence* $(x^k)_k$ *converges, as* $k \to \infty$*, to the centrality eigenvector when* $x_i^0 = \mathbb{1}$*.*

Proof. Thanks to **(H)**, the power iteration method is known to converge to x^* whenever $\langle x^*, x^0 \rangle \neq 0$, see [18]. Since x^* is a non-negative eigenvector, we can pick $i^* \in H^*$ such that $x_{i^*}^* > 0$. With this, it holds that $\langle x^*, x^0 \rangle = \sum_i x_i^* \geq x_{i^*}^* > 0$ because x^* is non-negative.

Katz Centrality. Eigenvector centrality may not be well-defined. Katz centrality addresses this problem by adding a predefined value $\beta > 0$ to each centrality. More formally, Katz centrality is given by $x_i^* = \alpha \sum_{j \in \mathcal{V}} A_{i,j} x_j + \beta$, where $\alpha > 0$ is some constant. Ultimately, this can be rewritten to $x^* = (I - \alpha A)^{-1} \mathbb{1}$, where β can be set to 1 without loss of generality.

Similarly to eigenvector centrality, Katz centrality is not obtained by solving the linear system $x^* = (I - \alpha A)^{-1} \mathbb{1}$ directly. Instead, the following iterative computation is used.

Theorem 2. *Assume that* $\|\alpha A\| < 1$*. Moreover, set* $x^0 := \mathbb{1}$ *and* $x^{k+1} := \alpha A x^k + \mathbb{1}$ *for all* $k \geq 0$*. Then,* $(x^k)_k$ *converges to the Katz centrality as* $k \to \infty$*.*

PageRank Centrality. A potential disadvantage of Katz centrality and eigenvector centrality is that nodes with high centrality propagate their authority to their peers. PageRank centrality accounts for this fact by dividing the propagated centrality by the number of outgoing edges of a node. More formally, the PageRank centrality is given by $x_i^* = \alpha \sum_{j \in \mathcal{V}} A_{i,j} \frac{x_j}{d_j} + \beta$, where d_j denotes the degree of node $j \in \mathcal{V}$. By removing nodes which have no outgoing edges, we may assume without loss of generality that $d_j \geq 1$ for all $j \in \mathcal{V}$. With this, it holds that $x^* = (I - \alpha D^{-1} A)^{-1} \mathbb{1}$, where D is the degree matrix, i.e., $D_{i,j} = d_i$ if $i = j$ and $D_{i,j} = 0$ otherwise.

The following holds true.

Theorem 3. *Define* $x^0 := \mathbb{1}$ *and* $x^{k+1} := \alpha D^{-1} A x^k + \mathbb{1}$ *for all* $k \geq 0$ *and some* $0 < \alpha < 1$*. Then,* $(x^k)_k$ *converges to the Pagerank centrality as* $k \to \infty$*.*

It is known that $D^{-1}A$ defines a discrete-time Markov chain with $\|D^{-1}A\|_\infty = 1$. The coefficient α is commonly known as damping factor and is usually set to 0.85.

2.2 Backward Differential Equivalence

We next provide a brief introduction to backward differential equivalence (BDE), a model reduction technique for dynamical systems [9,14]. For the benefit of presentation we restrict ourselves to linear dynamical systems in discrete time, even though the actual theory is stated for nonlinear continuous time dynamical systems.

Definition 3. *Fix the discrete time system $x^{k+1} = Ax^k$ with initial condition $x(0)$ and a partition $\mathcal{H}$ of the index set $\mathcal{V}$.*

- *Let $\mathcal{U}_\mathcal{H} \subseteq \mathbb{R}^\mathcal{V}$ be the linear subspace of vectors that are uniform on $\mathcal{H}$, that is, $\mathcal{U}_\mathcal{H} = \{x \in \mathbb{R}^\mathcal{V} \mid x_i = x_j, H \in \mathcal{H}, i, j \in H\}$.*
- *A partition $\mathcal{H}$ is BDE if $\mathcal{U}_\mathcal{H}$ is an invariant space of A, i.e., if $A(\mathcal{U}_\mathcal{H}) \subseteq \mathcal{U}_\mathcal{H}$.*

Example 1. Let us consider the dynamical system $x^{k+1} = Ax^k$ given by

$$\begin{aligned} x_1^{k+1} &= 0.5x_2^k + 0.5x_3^k \\ x_2^{k+1} &= 0.5x_2^k + 0.5x_1^k \\ x_3^{k+1} &= 0.5x_3^k + 0.5x_1^k \end{aligned} \tag{1}$$

The matrix underlying the above dynamical system is given by $A = ((0.0, 0.5, 0.5), (0.5, 0.5, 0.0), (0.5, 0.0, 0.5))^T$, where ^{T} denotes the transpose of a vector. It is not hard to see that $\mathcal{H} = \{\{1\}, \{2, 3\}\}$ is a BDE of $x^{k+1} = Ax^k$. ◊

We next present BDE reduction in the context of linear dynamical systems.

Definition 4. *Assume that $\mathcal{H} = \{H_1, \ldots, H_m\}$ is a BDE partition of the dynamical system $x^{k+1} = Ax^k$.*

- *For any $H \in \mathcal{H}$, fix some representative $i_H \in H$ of H and let $\hat{\mathcal{V}} = \{i_H \mid H \in \mathcal{H}\}$ denotes the set of representatives underlying $\mathcal{H}$.*
- *For any set $I \subseteq \mathcal{V}$, let $\mathbb{1}_I \in \mathbb{R}^\mathcal{V}$ be such that $\mathbb{1}_I(i) = 1$ if $i \in I$ and $\mathbb{1}_I(i) = 0$ otherwise.*
- *The reduced dynamical system $\hat{x}^{k+1} = \hat{A}\hat{x}^k$ with $\hat{x} \in \mathbb{R}^{\hat{\mathcal{V}}}$ arises from A in two steps. First, eliminate the equations of non-representative variables. Afterwards, replace any x_j by its representative, i.e., replace any x_j with $\hat{x}_{i_H}$ when $j \in H$ and $H \in \mathcal{H}$. In matrix language, this corresponds to $\hat{A} = (\mathbb{1}_{\{i_{H_1}\}}, \ldots, \mathbb{1}_{\{i_{H_m}\}})^T \cdot A \cdot (\mathbb{1}_{H_1}, \ldots, \mathbb{1}_{H_m})$.*

Example 2. Let us reduce (1) in the case when $i_{H_1} = 1$ and $i_{H_2} = 2$. We first remove the equation of x_3 because $H_2 = \{2, 3\}$. Afterwards, we replace any x_1 with $\hat{x}_1$, any x_2 with $\hat{x}_2$ and any x_3 with $\hat{x}_2$. From this we obtain $\hat{x}_1^{k+1} = \hat{x}_2$ and $\hat{x}_2^{k+1} = 0.5\hat{x}_1 + 0.5\hat{x}_2$. It is not hard to see that this corresponds to $\hat{x}^{k+1} = \hat{A}\hat{x}^k$ when $\hat{A}$ is computed using the algebraic expression from Definition 4:

$$\hat{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.0 & 1.0 \\ 0.5 & 0.5 \end{pmatrix}$$

◊

The relation between the original and the reduced dynamical system is as follows [12].

Theorem 4. *The following holds true.*

- *A partition $\mathcal{H}$ is a BDE of $x^{k+1} = Ax^k$ if and only if $x^0 \in \mathcal{U}_\mathcal{H}$ implies $x^k \in \mathcal{U}_\mathcal{H}$ for all $k \geq 0$.*
- *Let $\mathcal{H}$ be a BDE of $x^{k+1} = Ax^k$. Then, if $\hat{x}^0_{i_H} = x^0_{i_H}$ for all $H \in \mathcal{H}$ and $\hat{x}$ denotes the solution of the BDE reduction $\hat{x}^{k+1} = \hat{A}\hat{x}^k$ subject to $\hat{x}^0$, it holds that $\hat{x}^k_{i_H} = x^k_{i_H}$ for all $k \geq 0$ and $H \in \mathcal{H}$.*

Theorem 4 ensures that the original dynamical solution can be obtained by solving the reduced dynamical system whenever the initial condition x^0 is uniform on $\mathcal{H}$, i.e., when $x^0 \in \mathcal{U}_\mathcal{H}$.

Proof. See [9,37, Theorem 3].

Example 3. If $x^0 = (0.2, 0.4, 0.4)$, then the solution of (1) satisfies $x^k_2 = x^k_3$ for all $k \geq 0$. Moreover, if $\hat{x}^0_1 = 0.2$, $\hat{x}^0_2 = 0.4$, $\hat{x}^{k+1}_1 = \hat{x}^k_2$ and $\hat{x}^{k+1}_2 = 0.5\hat{x}^k_1 + 0.5\hat{x}^k_2$, it holds that $\hat{x}^k_1 = x^k_1$ and $\hat{x}^k_2 = x^k_2 = x^k_3$ for all $k \geq 0$. ◊

The following result ensures that there exists a unique coarsest BDE partition $\mathcal{H}$ that yields the best possible reduction.

Proposition 1. *For a given dynamical system $x^{k+1} = Ax^k$, there exists a coarsest BDE partition $\mathcal{H}$. That is, for any BDE partition $\mathcal{H}'$ of $x^{k+1} = Ax^k$, it holds that $\mathcal{H}'$ refines $\mathcal{H}$.*

Proof. See [9, Theorem 3].

The coarsest BDE partition can be computed efficiently [9,14] as stated next.

Theorem 5. *There exists a partition refinement algorithm that needs $\mathcal{O}(|A| \cdot \log(|\mathcal{V}|))$ steps to compute the coarsest BDE partition of $x^{k+1} = Ax^k$, where $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ and $|A|$ denotes the number of non-zero entries in A.*

Proof. See [9, Theorem 3].

Remark 1. The section discusses BDE in the context of linear dynamical systems. We wish to point out, however, that [12] captures nonlinear dynamical systems of the form $x^{k+1} = F(x^k)$. In particular, the partition refinement algorithm from [14] generalizes to the case when F is given in terms of multi-variate polynomials and enjoys a polynomial time and space complexity in the number of monomials present in F.

2.3 Stochastic Lumpability

We next provide an account on lumpability of Markov chains [7] and relate it to BDE.

Definition 5. *Given a set of nodes $\mathcal{V}$, we define the following.*

- *A transition matrix $P \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ satisfies $p_{i,j} \geq 0$ and $\sum_k p_{i,k} = 1$ for all $i, j \in \mathcal{V}$.*
- *For an initial probability distribution π^0 on $\mathcal{V}$, the transient probabilities of the discrete time Markov chain (DTMC) induced by P are given by $(\pi^{k+1})^T = (\pi^k)^T P$.*
- *A partition $\mathcal{H}$ is an exactly lumpable partition of P when, for all $k > 0$, $\pi^k \in \mathcal{U}_\mathcal{H}$ for all initial probability distributions $\pi^0 \in \mathcal{U}_\mathcal{H}$ and $t \geq 0$.*

The entry $p_{i,j}$ corresponds to the transition probability from state i into state j. The original matrix from Example 2 is an example of a transition matrix.

The next result is well-known in the area of Markov chains [7].

Theorem 6. *A partition $\mathcal{H}$ is an exactly lumpable partition of a transition matrix P if and only if, for any $H, H' \in \mathcal{H}$ and $i, j \in H$ it holds that $\sum_{k\in H'} P_{k,i} = \sum_{k\in H'} P_{k,j}$.*

Theorem 6 states essentially that $\mathcal{H}$ is an exactly lumpable partition whenever the cumulative transition probabilities from block H' into any two nodes of block H coincide.

Similarly to BDE, it is possible to define a reduced transition matrix $\hat{P}$ which underlies an exactly lumpable partition $\mathcal{H}$ of P.

Theorem 7. *For an exactly lumpable partition $\mathcal{H}$ of P, set $\hat{P}_{i_H,i_{H'}} := \sum_{k\in H} P_{k,i_{H'}}$ for all $H, H' \in \mathcal{H}$. Then, $\hat{P}$ defines a lumped DTMC with states $\{i_H \mid H \in \mathcal{H}\}$ whose transient probabilities $\hat{\pi}$ satisfy $\hat{\pi}^k_{i_H} = \pi^k_{i_H}$ for all $H \in \mathcal{H}$ and $k \geq 0$ if the initial probability distribution obeys $\pi^0 \in \mathcal{U}_\mathcal{H}$.*

It can be proven that BDE coincides with exact lumpability on the domain of DTMCs.

Theorem 8. *Let P be a transition matrix and $\mathcal{H}$ a partition of $\mathcal{V}$. Then, $\mathcal{H}$ is an exactly lumpable partition of P if and only if $\mathcal{H}$ is a BDE of P^T.*

Proof. In [12] it has been shown that this is true for continuous time Markov chains (CTMCs). Any CTMC can be turned into a DTMC (and vice versa) while preserving many important properties, including lumpability [7]. This shows the statement.

Remark 2. The transpose in Theorem 8 is due to the fact that probabilities are commonly denoted as row vectors instead of column vectors. Because of this, the dynamical system of π is given by $(\pi^{k+1})^T = (\pi^k)^T P$, while that of x is given by $x^{k+1} = Ax^k$.

There are established algorithms for the efficient computation of the coarsest exactly lumpable partition of a Markov chain [16,38]. In fact, the partition refinement algorithm from [9,14] can be seen as an extension of [16,38] to general dynamical systems. While the specialized algorithm [16,22] has been reported to outperform the more general algorithm [9], the complexity bounds of both algorithms coincide for Markov chains [9].

2.4 Exact Role Assignment

We start by giving the definition of exact role assignment [26,40].

Definition 6. *Given a symmetric adjacency matrix $A \in \{0,1\}^{\mathcal{V}\times\mathcal{V}}$ and $i \in \mathcal{V}$, let $N^+(i)$ denote the out-neighbors of i, that is $N^+(i) := \{j \in \mathcal{V} \mid (i,j) \in A\}$. A surjective mapping $r : \mathcal{V} \to \hat{\mathcal{V}}$ with $\hat{\mathcal{V}} \subseteq \mathcal{V}$ is an exact role assignment if, for all $i, j \in \mathcal{V}$, $r(i) = r(j)$ implies $r(N^+(i)) = r(N^+(j))$.*

The following result allows one to express exact role assignment as a bisimulation.

Theorem 9. *Given a symmetric adjacency matrix $A \in \{0,1\}^{\mathcal{V}\times\mathcal{V}}$ and a partition $\mathcal{H}$ of $\mathcal{V}$, set $r(i) = i_H$ when $i \in H$ for $H \in \mathcal{H}$.*

- *r is an exact role assignment if and only if for all $H, H' \in \mathcal{H}$ and $i, j \in H$, it holds that $\sum_{k\in H'} A_{i,k} = \sum_{k\in H'} A_{j,k}$.*
- *It holds that $\sum_{k\in H'} A_{i,k} = \sum_{k\in H'} A_{j,k}$ if and only if $\sum_{k\in H'} A_{k,i} = \sum_{k\in H'} A_{k,j}$ is valid.*

Remark 3. The sum criterion of the exact role assignment resembles the exact lumpability from Theorem 6 because $P = A^T$, see Theorem 8 and Remark 2. Note, however, that A is not a stochastic matrix, as required by exact lumpability.

3 Centrality and Model Reduction

Clustering techniques such as minimal cut [1], normalized cut [35] or k-means [39] clustering have been studied substantially in the past. In [30] normalized cut has been related to the eigenvector corresponding to the second largest eigenvalue of the transition matrix P of the normalized Laplace matrix.

The following results shows that BDE is a natural generalization of exact role assignment (and exact lumpability, as has been already observed in [12]).

Theorem 10. *Let $A \in \mathbb{R}^{\mathcal{V}\times\mathcal{V}}$ (in particular, A may have negative entries) and $\mathcal{H}$ a partition of $\mathcal{V}$.*

(i) $\mathcal{H}$ is a BDE of $x^{k+1} = Ax^k$ if and only if, for all $H, H' \in \mathcal{H}$ and $i, j \in H$, it holds that $\sum_{k\in H'} A_{i,k} = \sum_{k\in H'} A_{j,k}$.

(ii) In the case when A is an adjacency matrix of an undirected graph, $\mathcal{H}$ is a BDE of $x^{k+1} = Ax^k$ if and only if, for all $H, H' \in \mathcal{H}$ and $i, j \in H$, it holds that $\sum_{k\in H'} A_{i,k} = \sum_{k\in H'} A_{j,k}$ and $\sum_{k\in H'} A_{k,i} = \sum_{k\in H'} A_{k,j}$.

(iii) In the case when A is an adjacency matrix of an undirected graph, $\mathcal{H}$ is a BDE of $x^{k+1} = Ax^k$ if and only if $\mathcal{H}$ is an exact role assignment.

Proof. Since $\mathcal{H}$ is a BDE of A, it holds that $A \cdot \mathbb{1}_{H'} \in \mathcal{U}_{\mathcal{H}}$ which in turn implies $\mathbb{1}_{\{i\}} \cdot A \cdot \mathbb{1}_{H'} = \mathbb{1}_{\{j\}} \cdot A \cdot \mathbb{1}_{H'}$. This yields the first claim. The second claim follows trivially thanks to $A = A^T$. The third claim is a direct consequence of Theorem 9

Additionally to the graph theoretical characterization of BDE from Theorem 10, we next show that any two nodes of a BDE block enjoy identical Katz, pagerank and eigenvector centrality.

Definition 7. *Fix a non-negative matrix A and assume that the diagonal of the degree matrix D, given by $d_{ii} = \sum_{j \neq i} A_{i,j}$, is positive.*

- *Let $\mathcal{H}$ be a BDE of $x^{k+1} = Ax^k$ and let $\hat{A}$ denote the corresponding BDE reduction. Then, the reduced power iteration sequence $(\hat{x}^k)_k$ is given by $\hat{x}^0 := \hat{\mathbb{1}}$ and $\hat{x}^{k+1} = \hat{A}\hat{x}^k$ for all $k \geq 0$, where $\hat{\mathbb{1}}_{i_H} = 1$ for all $H \in \mathcal{H}$.*
- *Let $\mathcal{H}$ be a BDE of $x^{k+1} = Ax^k$ and let $\hat{A}$ denote the corresponding BDE reduction. Then, for any $0 < \alpha < \|A\|$, the reduced Katz sequence is given by $\hat{x}^0 := \hat{\mathbb{1}}$ and $\hat{x}^{k+1} := \alpha\hat{A}\hat{x}^k + \hat{\mathbb{1}}$ for all $k \geq 0$.*
- *Let $\mathcal{H}$ be a BDE of $x^{k+1} = D^{-1}Ax^k$ and let $\hat{A}$ denote the corresponding BDE reduction. Then, for any $0 < \alpha < 1$, the reduced Pagerank sequence is given by $\hat{x}^0 := \hat{\mathbb{1}}$ and $\hat{x}^{k+1} := \alpha\hat{A}\hat{x}^k + \hat{\mathbb{1}}$ for all $k \geq 0$.*

Similarly to the case of ODE systems, the original sequences can be obtained from the reduced ones.

Theorem 11. *Fix a non-negative matrix A, assume that the diagonal of the degree matrix D is positive and let $\mathcal{H}$ be a BDE of $x^{k+1} = Ax^k$.*

(i) $\mathcal{H}$ is also a BDE of $x^{k+1} = D^{-1}Ax^k$, where D denotes the degree matrix of A.

*(ii) For any centrality measure from Definition 7, $x^* \in \mathcal{U}_{\mathcal{H}}$ and the corresponding reduced sequence $(\hat{x}^k)_k$ converges to $\hat{x}^* \in \mathbb{R}^{\hat{V}}$, where $\hat{x}^*_{i_H} = x^*_{i_H}$ for all $H \in \mathcal{H}$ (in the case of eigenvector centrality, we additionally require **(H)**).*

Proof. Note that Theorem 10 implies $\sum_{k \in H'} A_{i,k} = \sum_{k \in H'} A_{j,k}$ for all $H, H' \in \mathcal{H}$ and $i, j \in H$. This yields $d_i = \sum_{H' \in \mathcal{H}} \sum_{k \in H'} A_{i,k} = \sum_{H' \in \mathcal{H}} \sum_{k \in H'} A_{j,k} = d_j$. With this, we infer that $\mathcal{H}$ is a BDE of $x^{k+1} = D^{-1}Ax^k$. In the following we prove the statements concerning reduced sequences. Eigenvector centrality: The fact that $\mathcal{H}$ is a BDE of $x^{k+1} = Ax^k$ ensures that $A(\mathcal{U}_{\mathcal{H}}) \subseteq \mathcal{U}_{\mathcal{H}}$. Hence, the power iteration sequence remains in $\mathcal{U}_{\mathcal{H}}$ and the statement follows from Theorem 1. Katz centrality: Similarly to eigenvector centrality, we know that $A(\mathcal{U}_{\mathcal{H}}) \subseteq \mathcal{U}_{\mathcal{H}}$. Since $\mathbb{1} \in \mathcal{U}_{\mathcal{H}}$, we obtain $x^{k+1} = \alpha Ax^k + \mathbb{1} \in \mathcal{U}_{\mathcal{H}}$ whenever $x^k \in \mathcal{U}_{\mathcal{H}}$ and Theorem 2 yields the claim. Pagerank centrality: Since $\mathcal{H}$ is a BDE of $x^{k+1} = D^{-1}Ax^k$, we conclude that $D^{-1}A(\mathcal{U}_{\mathcal{H}}) \subseteq \mathcal{U}_{\mathcal{H}}$. Arguing as in the case of Katz centrality and invoking Theorem 3 instead of Theorem 2, we obtain the claim.

Theorem 11 states that members of the same group have the same centrality measure regardless whether Katz, pagerank or eigenvector centrality is used. The computation of the centrality measures via the reduced sequence is outlined in Algorithm 1. Note also that (i) implies that the coarsest BDE $\mathcal{H}$ of A is the coarsest (ordinarily) lumpable partition of the transition matrix $D^{-1}A$, see [7].

Require: Non-negative matrix A, numerical threshold $\delta > 0$.
procedure ComputeCentralityViaRedSequence(A)
$\eta \leftarrow \infty$
$\hat{x}^{\text{old}} \leftarrow \hat{\mathbb{1}}$
$\hat{A} \leftarrow ComputeCoarsestReduction(A)$ ▷ See Theorem 5
while $\eta \geq \delta$ **do**
$\hat{x} \leftarrow ComputeNextSequenceElement(\hat{A}, \hat{x}^{\text{old}})$
$\eta \leftarrow \|\hat{x} - \hat{x}^{\text{old}}\|$
$\hat{x}^{\text{old}} \leftarrow \hat{x}$
end while
$x^* \leftarrow Expand(\hat{x})$ ▷ $x^*_i := \hat{x}^*_{i_H}$ for all $H \in \mathcal{H}, i \in H$
return x^*
end procedure

Fig. 1. Centrality computation via reduced sequence.

Hence, it is in principle possible to apply the specialized partition refinement algorithms for Markov chains [16,22,38] instead of the one for general dynamical systems [9]. We wish to stress, however, that the proof of this fact requires the notion of BDE and Theorem 11.

Theorem 11 yields the following.

Corollary 1. *Exact role assignment yields Katz, pagerank and eigenvector centrality.*

As pointed out earlier, eigenvectors of adjacency matrices are crucial for clustering. The eigenvectors of the BDE reduction $\hat{A}$ will provide only a proper subset of all eigenvectors in general. In particular, it is not clear how to check whether the subset contains the k largest eigenvector without considering the original matrix.

We wish to stress the following.

Remark 4. While it is known that exact role assignment preserves certain algebraic properties such as eigenvalues [26], we are not aware of any result that would establish that exact role assignment preserves Katz, pagerank and eigenvector centrality.

4 Experimental Results

In this section we present the results of our experimental evaluation on some real world case studies [15,29]. We measure the performance of our approach in terms of model reduction ratio.

Implementation and Environment. The input is the adjacency matrix A of a graph $G = (\mathcal{V}, E)$. In Theorem 5 an efficient algorithm is presented to compute the coarsest BDE, such algorithm is implemented in the tool ERODE (Evaluation and Reduction of ODEs) [13]. The experimental evaluation hereby presented

has been performed using a Matlab prototype linked to ERODE which was used to compute the reduced models. We present results on *directed* and *undirected* graphs. All the experiments ran on a machine with an Intel Xeon E7-4830v4, with a 64-bits architecture at 2 GHz, 14 cores, 112 CPUs and 500 GB of RAM.

The Instances. In order to provide some real-world case studies we ran our proposed reduction technique on some networks obtained from the SNAP (Stanford Large Network Dataset Collection) and the Florida Sparse Matrix collection repositories.

We first present the *undirected graphs* instances:

- *GD06-Theory*: this is a hierarchical network with 3 levels, proposed in a graph design contest as an "artificially symmetric" one. The main node is connected to 9 children nodes, each of which is connected to 9 children nodes. With this design there are exactly 3 classes if we partition with respect to the centrality measures, because all of the grandchildren will have the same score (the lowest). Intuitively, this shows that a natural interpretation of the blocks is obtained from our proposed methodology. We remark that, although it is hard to find real-world examples which are *exactly* symmetric like this particular instance, Table 2 shows that we achieve good reductions on real-world networks.
- *Yeast protein interaction network*: this network was developed by Barabasi et al. [21] in order to study the interaction between proteins in yeast.
- *Collaboration networks*: we have four different collaboration networks that belong in this category. In our results we present a collaboration network in the field of general relativity research from Arxiv [28], the Erdős collaboration network [4] and two different co-authorship networks from Citeseer. It has been shown in [3,34] that this type of collaboration networks obtained from real-world problems tend to have an hierarchical structure; this confirms that we obtain good reductions with our notion of reduction on hierarchical networks.
- *Autonomous Systems*: these networks are obtained from routing networks, these networks reduce well and include one autonomous system from the SNAP repository which contains 733 different daily snapshots of the same graph [27]. In Table 2 we present only one of those instances as they have a similar rate of reduction to one another. This is expected as this graph is a stable network (i.e., there is no drastic change in the network from one day to another). This same concept applies to the Oregon routing network instance.
- *Enron email network*: this is a classic benchmark of a network obtained by tracking the email interactions between the members of the Enron Corporation [23].
- *YouTube*: this is a network of interactions between YouTube channels [31].

The aforementioned case studies are the subset of the datasets on which our technique proved to be the most efficient on in terms of reduction. We present the results with respect to the reduction ratio. The reduction ratio is defined as the ratio of the size of the reduced model and the size of the original model. The

Table 1. Case studies results

Undirected Graphs			
Instance	Original size	Reduced size	Reduction ratio
GD06-Theory	102	3	2.94%
Yeast protein interaction network	1871	1091	58.31%
Collaborations in General Relativity	5243	3394	64.73%
Erdős collaboration network	5535	1902	34.36%
Autonomous system (SNAP)	6475	3691	57.00%
Oregon routing network	10671	5484	51.39%
Autonomous system (Florida)	22964	11935	51.97%
Enron email network	36693	20418	55.65%
Dictionary	39328	26994	68.64%
Caida routers	192245	150463	78.27%
Citeseer coauthorship network	227321	155593	68.45%
Citeseer copaper network	434103	150316	34.63%
YouTube	1134891	684011	60.27%

results are presented in Table 2. Each row represents an instance from the list presented in the last paragraph. We show its original size, its reduced size and the reduction ratio.

We obtain considerable reductions for networks of different sizes. We wish to stress again that the reduction is *exact*, meaning that two nodes are in the same block if and only if they exactly have the same eigenvector centrality score. This is the reason behind the fact that the best reduction is obtained from the instance *GD-06*: as explained earlier, this instance was built artificially and its property of having a symmetric defined three-level hierarchical structure induces a partition which is composed by three blocks, one per level. In real world networks it is rare that two or more nodes share exactly the same centrality score because of the fact that real world networks do not have a particularly symmetrical hierarchical structure due to the inhomogeneous nature of interactions between peers. Despite of this, we are able to produce good reductions, particularly in networks that arise from academic collaborations. The intuitive reason is that these types of collaboration tend to have a more regular and symmetric hierarchical structure; therefore the chances of two nodes having exactly the same eigenvector centrality scores are higher.

Table 2. Case studies results

Directed Graphs			
Instance	Original size	Reduced size	Reduction ratio
Glossary	73	41	56.16%
Graph Design '96	112	6	5.36%
PhDs in computer science	1883	225	11.95%
Kohonen citation network	4471	766	17.13%
EPA web pages	4733	598	12.63%
Gnutella p2p network	6302	2208	35.04%
Wikipedia who-votes-on-whom	8299	4216	50.80%
EVA corporate inter-relationships	8498	215	2.53%
California web search	9665	1817	18.80%
Stanford CS web	9915	3657	36.88%
Gnutella p2p network (I)	10880	4340	39.89%
Gnutella p2p network (II)	26519	6741	25.42%
Enron email traffic	69245	7437	10.74%
Epinions trust network	75889	41055	54.10%
Slashdot social network	82169	57561	70.05%
Stanford web graph	281905	129335	45.88%
CNR web crawl	325558	85419	26.24%
Notre Dame web graph	325730	49952	15.34%
Berkely.edu + stanford.edu web	685252	292492	42.68%
Flickr web crawl	820879	370145	45.09%
.eu domain web crawl	862665	341687	39.60%
Google web graph	916429	354624	38.70%
.in domain web crawl	1382909	333283	24.10%
Wikipedia pages	1634990	1116472	68.29%

Similarly, we provide a set of *directed graphs* case studies:

- *Academic instances*: *PhD in Computer Science* and *Kohonen citation network* arise from academic real-world examples. The first instance describes a network where an edge from node i to node j means that i is a PhD student of j while the latter describes a network of citations.
- *Web infrastructures*: different instances of web infrastructure are presented. *EPA web pages*, *Gnutella p2p network*, *Stanford Computer Science*, *Stanford web graph*, *CNR web*, *Notre Dame web*, *Berkley and Stanford domains*, *European domain web*, *Google web*, *Indian domain web*, *Flickr web* and *Wikipedia* all fall in this category. The main underlying theme of all these instances is that are generated from world wide web problems. Nodes represent web pages while directed edges from one node to another are hyperlinks. Some instances

appear more than once in Table 2 (for example, Gnutella p2p network appears multiple times), as we present the results on the same network infrastructure on different days. Differently from the undirected instances, these directed instances are not stable and they present differences in size and structure on different days.

– *Enron*: differently from the instance provided in its undirected version, here we link together two nodes with the meaning of i has sent a mail to j.
– *Social Networks*: we present some benchmarks on trust networks and social networks. Such instances are the following: *Wikipedia who-votes-on-whom*, *Epinions trust network* and *Slashdot social network*. The first is a network in which an edge has a source node i and a target node j if i voted for j. The *Epinions trust network* describes the relationship between users of epinions.com. This social network is a general consumer review website where members can decide whether to *trust* the reviews of other members. The trust relationships form the web of trust which is then combined with review ratings to determine which reviews are shown to a user. Last, slashdot.org is a technology-related news website that features user-submitted and editor-evaluated news. In 2002 Slashdot introduced the Slashdot Zoo which allowed users to tag each other as friend or foes. The presented instance captures those relationships obtained in February 2009.

5 Conclusions and Future Work

In this paper we have related network centrality to differential equivalence, a model reduction technique that generalizes stochastic lumpability. We have shown that differential equivalence coincides with the exact role assignment on undirected graphs and that differential equivalent nodes have the same Katz, pagerank and eigenvector centrality. The relevance of the result was demonstrated by efficiently computing substantial reductions of published real-world networks, including protein interaction networks, scientific collaboration networks, routing networks, email networks and networks capturing YouTube.

Future work will focus on the development of approximate notions of differential equivalence and the study of already established notions of ε-lumpability and near-lumpability of Markov chains. Thanks to the theory established in this paper, this will naturally lead to approximate exact role assignments. The need for an approximate version of exact role assignments is motivated by the fact that nodes on the "periphery" of a network are often distinguished by differential equivalence while featuring almost identical centrality measures.

References

1. Arora, S., Rao, S., Vazirani, U.V.: Expander flows, geometric embeddings and graph partitioning. J. ACM **56**(2), 5:1–5:37 (2009)
2. Baier, C., Engelen, B., Majster-Cederbaum, M.E.: Deciding bisimilarity and similarity for probabilistic processes. J. Comput. Syst. Sci. **60**(1), 187–231 (2000)

3. Barabasi, A.-L., Dezso, Z., Regan, E., Yook, S.-H., Oltvai, Z.: Scale-free and hierarchical structures in complex networks, vol. 661, no. 1 (2003)
4. Batagelj, V., Mrvar, A.: Pajek datasets, June 2006. http://vlado.fmf.uni-lj.si/pub/networks/data/
5. Bobbio, A., Gribaudo, M., Telek, M.: Analysis of large scale interacting systems by mean field method. In: QEST, pp. 215–224 (2008)
6. Bortolussi, L., Gast, N.: Mean field approximation of uncertain stochastic models. In: DSN, pp. 287–298 (2016)
7. Buchholz, P.: Exact and ordinary lumpability in finite Markov chains. J. Appl. Probab. **31**, 59–74 (1994)
8. Cardelli, L., Csikász-Nagy, A., Dalchau, N., Tribastone, M., Tschaikowski, M.: Noise Reduction in complex biological switches. Sci. Rep. (2016)
9. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Efficient syntax-driven lumping of differential equations. In: Chechik, M., Raskin, J.-F. (eds.) TACAS 2016. LNCS, vol. 9636, pp. 93–111. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49674-9_6
10. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Forward and backward bisimulations for chemical reaction networks. In: CONCUR, pp. 226–239 (2015)
11. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Comparing chemical reaction networks: a categorical and algorithmic perspective. In: LICS, pp. 485–494 (2016)
12. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Symbolic computation of differential equivalences. In: 43rd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL) (2016)
13. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: ERODE: a tool for the evaluation and reduction of ordinary differential equations. In: Legay, A., Margaria, T. (eds.) TACAS 2017. LNCS, vol. 10206, pp. 310–328. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-662-54580-5_19
14. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Maximal aggregation of polynomial dynamical systems. PNAS **114**(38), 10029–10034 (2017)
15. Davis, T.A., Yifan, H.: The university of florida sparse matrix collection. ACM Trans. Math. Softw. **38**(1), 1:1–1:25 (2011)
16. Derisavi, S., Hermanns, H., Sanders, W.H.: Optimal state-space lumping in Markov chains. Inf. Process. Lett. **87**(6), 309–315 (2003)
17. Feret, J., Danos, V., Krivine, J., Harmer, R., Fontana, W.: Internal coarse-graining of molecular systems. PNAS **106**(16), 6453–6458 (2009)
18. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
19. Iacobelli, G., Tribastone, M.: Lumpability of fluid models with heterogeneous agent types. In: DSN, pp. 1–11 (2013)
20. Iacobelli, G., Tribastone, M., Vandin, A.: Differential bisimulation for a Markovian process algebra. In: Italiano, G.F., Pighizzini, G., Sannella, D.T. (eds.) MFCS 2015. LNCS, vol. 9234, pp. 293–306. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48057-1_23
21. Jeong, H., Mason, S.P., Barabasi, A.-L., Oltvai, Z.N.: Lethality and centrality in protein networks. Nature **411**(6833), 41–42 (2001)
22. Katoen, J.-P., Zapreev, I.S., Hahn, E.M., Hermanns, H., Jansen, D.N.: The ins and outs of the probabilistic model checker MRMC. Perform. Eval. **68**(2), 90–104 (2011)

23. Klimt, B., Yang, Y.: The enron corpus: a new dataset for email classification research. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 217–226. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30115-8_22
24. Kowal, M., Tschaikowski, M., Tribastone, M., Schaefer, I.: Scaling size and parameter spaces in variability-aware software performance models (T). In: ASE, pp. 407–417 (2015)
25. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. In: POPL, pp. 344–352 (1989)
26. Lerner, J.: Role assignments. In: Brandes, U., Erlebach, T. (eds.) Network Analysis: Methodological Foundations. LNCS, vol. 3418, pp. 216–252. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-31955-9_9
27. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: KDD 2005, pp. 177–187 (2005)
28. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: densification and shrinking diameters. ACM Trans. Knowl. Discov. Data **1**(1), 2 (2007)
29. Leskovec, J., Krevl, A.: SNAP datasets: Stanford large network dataset collection, June 2014. http://snap.stanford.edu/data
30. Meila, M., Shi, J.: Learning segmentation by random walks. In: NIPS 2000, Denver, pp. 873–879 (2000)
31. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC 2007), San Diego, October 2007
32. Newman, M.: Networks: An Introduction. Oxford University Press Inc., New York (2010)
33. Paige, R., Tarjan, R.E.: Three partition refinement algorithms. SIAM J. Comput. **16**(6), 973–989 (1987)
34. Ravasz, E., Barabási, A.-L.: Hierarchical organization in complex networks. Phys. Rev. E **67**, 026112 (2003)
35. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)
36. Tschaikowski, M., Tribastone, M.: Exact fluid lumpability for markovian process algebra. In: CONCUR, pp. 380–394 (2012)
37. Tschaikowski, M., Tribastone, M.: Approximate reduction of heterogenous nonlinear models with differential hulls. IEEE TAC **61**(4), 1099–1104 (2016)
38. Valmari, A., Franceschinis, G.: Simple $O(m \log n)$ time Markov chain lumping. In: Esparza, J., Majumdar, R. (eds.) TACAS 2010. LNCS, vol. 6015, pp. 38–52. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12002-2_4
39. Vinnikov, A., Shalev-Shwartz, S.: K-means recovers ICA filters when independent components are sparse. In: ICML 2014, Beijing, 21–26 June 2014, pp. 712–720 (2014)
40. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, vol. 8. Cambridge University Press, Cambridge (1994)