New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

OPEN ACCESS

RECEIVED
16 August 2015

REVISED
9 November 2015

ACCEPTED FOR PUBLICATION
23 November 2015

PUBLISHED
24 December 2015

Minimum-cost control of complex networks

Guoqi Li[1,2], Wuhua Hu[2], Gaoxi Xiao[2], Lei Deng[1], Pei Tang[1], Jing Pei[1] and Luping Shi[1]

[1] Center for Brain Inspired Computing Research, Department of Precision Instrument, Tsinghua University, 100084, People's Republic of China
[2] School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore

E-mail: liguoqi@tsinghua.edu.cn, egxxiao@ntu.edu.sg and lpshi@mail.tsinghua.edu.cn

Keywords: complex networks, minimum energy cost, projected gradient method

Abstract

Finding the solution for driving a complex network at the minimum energy cost with a given number of controllers, known as the minimum-cost control problem, is critically important but remains largely open. We propose a projected gradient method to tackle this problem, which works efficiently in both synthetic and real-life networks. The study is then extended to the case where each controller can only be connected to a single network node to have the lowest connection complexity. We obtain the interesting insight that such connections basically avoid high-degree nodes of the network, which is in resonance with recent observations on controllability of complex networks. Our results provide the first technical path to enabling minimum-cost control of complex networks, and contribute new insights to locating the key nodes from a minimum-cost control perspective.

1. Introduction

The control of network systems is an important problem with wide applications. Examples include stabilizing or driving social systems [1, 2], management of socio-ecological systems [3, 4] and cyber-physical systems [5, 6], coordination of multi-agent systems [7, 8] such as mobile sensor systems [9, 10] and robotic systems [11, 12], prevention of cascading failures in interdependent systems such as smart grids and telecom systems [13, 14], etc. Significant studies have mainly focused on networks with N-dimensional linear time-invariant (LTI) dynamics [15–17]:

$$\dot{x}(t) = Ax(t) + Bu(t), \;\; x(0) = x_0, \tag{1}$$

where $x(t) = [x_1(t),\ldots, x_N(t)]^{\mathrm{T}}$ is the state vector of N nodes at time t with the initial state being x_0, $u(t) = [u_1(t),\ldots, u_M(t)]^{\mathrm{T}}$ is the time dependent input vector of external signals with $M \;\; (M \leqslant N)$ being the number of controllers where the same $u_i(t)$ may drive multiple nodes. The $N \times N$ matrix A is the network's adjacency matrix; i.e., $a_{ij} = 1$ if there exists a link from node i to node j; otherwise $a_{ij} = 0$. The $N \times M$ matrix B is the input matrix, where B_{im} is non-zero if controller m is connected to node i and zero otherwise. Without loss of generality, it is assumed that B has a full column rank to avoid the redundant inputs.

Two central issues in controlling LTI systems are (i) finding the minimum number of driver nodes connected to external controllers to ensure the system's controllability and (ii) for a given number of controllers, finding the solution of driving the system to any predefined state with the minimum cost. The former issue is known as the *network controllability* problem, and the latter the *minimum-cost control* problem. While a significant breakthrough has been made on the network controllability problem [16–20], and more recently on the relationship between network controllability and control cost [21, 22], the *minimum-cost control problem* largely remains as an open issue [23].

2. Minimum-cost control

The minimum-cost control problem can be modelled as driving the state vector to the origin in the time interval $[0, \; t_{\mathrm{f}}]$ with the minimum energy cost [23]

$$\mathcal{E}\left(t_{\mathrm{f}}\right) = \min_{u(t),B} \mathbb{E}\left[\int_0^{t_{\mathrm{f}}} \|u(t)\|^2 \mathrm{d}t\right] \tag{2}$$

under the constraint that the network (A, B) is controllable while B maintains a full rank subject to a normalization condition. The operator $\mathbb{E}[\bullet]$ takes the expectation of the argument over all realizations of the random initial state. Note that $u(t)$ for t from 0 to t_{f} and B are the decision variables to be determined by the optimization process, and their optimal solutions give the optimal input signals and their optimal connections with the network nodes.

Given an input matrix B such that (A, B) is controllable, existing results show that the minimum-cost control is achieved when

$$u(t) = -B^{\mathrm{T}}\mathrm{e}^{A^{\mathrm{T}}\left(t_{\mathrm{f}}-t\right)}W_B^{-1}x_{\mathrm{f}}, \tag{3}$$

where $W_B = \int_0^{t_{\mathrm{f}}} \mathrm{e}^{At}BB^{\mathrm{T}}\mathrm{e}^{A^{\mathrm{T}}t}\mathrm{d}t$ is the controllability Grammian matrix [24, 25], and $x_{\mathrm{f}} = \mathrm{e}^{At_{\mathrm{f}}}x_0$ is the final state of the network in the absence of $u(t)$. By fixing the quadratic sum of all elements of B and substituting (3) into (2), the *minimum-cost control* problem can be rewritten as

$$\begin{aligned} &\min \quad \mathrm{tr}\left[W_B^{-1}X_{\mathrm{f}}\right] \\ &\text{s.t.} \quad \mathrm{tr}\left(B^{\mathrm{T}}B\right) - M - \epsilon = 0 \end{aligned} \tag{4}$$

where $\mathrm{tr}[\bullet]$ denotes a matrix trace and X_{f} is a constant matrix given by $X_{\mathrm{f}} = \mathcal{E}[x_{\mathrm{f}}x_{\mathrm{f}}^{\mathrm{T}}] = \mathrm{e}^{At_{\mathrm{f}}}X_0\mathrm{e}^{A^{\mathrm{T}}t_{\mathrm{f}}}$, $E(B) \triangleq \mathrm{tr}[W_B^{-1}X_{\mathrm{f}}]$ is the defined energy cost function to be minimized. We also define a norm function $N(B) \triangleq (\mathrm{tr}(B^{\mathrm{T}}B) - M)^2$ to associate with the equality constraint expressing the normalization condition on B, where ϵ is a positive constant to ensure that $N(B)$ is non-zero on the surface $\mathrm{tr}(B^{\mathrm{T}}\mathrm{B}) = M + \epsilon$. The optimization has only the matrix B as its decision variable, but is hard to solve analytically or numerically because of the complicated inverse operation in the cost function. This has kept the minimum-cost control of complex networks, with its high practical and theoretical importance, an open problem.

Remark 1. There are different ways to normalize the matrix B, each corresponding to a set of distinctive boundary conditions. In this paper, we adopt the constraint that $\mathrm{tr}(B^{\mathrm{T}}B) = M + \epsilon$, which reflects the boundary conditions of the energy profile of the input matrix $B \in R^{N \times M}$ with physical implications. Specifically, it ensures that the quadratic sum of all the elements of the input matrix B has a fixed value. As shown later in equation (14), we project $-\frac{\partial E(B)}{\partial B}B$ onto $\frac{\partial N(B)}{\partial B}^{\perp}$ using the operator $\left(I - \left\{\frac{\partial N(B)}{\partial B}\right\}\left\{\frac{\partial N(B)}{\partial B}\right\}^{+}\right)$, which requires $N(B)$ to be a non-zero vector on the surface $\mathrm{tr}(B^{\mathrm{T}}B) = M$. This is the reason why a non-zero ϵ is needed. Since ϵ can be a constant of arbitrary value, it can certainly be a sufficiently small positive number.

We propose an efficient *projected gradient method* (PGM) to solve the optimization problem defined above. The main idea is to analytically derive the derivatives of the functions $N(B)$ and $E(B)$ with respect to B, and then project the negative of the gradient of $E(B)$ onto the sphere surface $\mathrm{tr}(B^{\mathrm{T}}B) = M + \epsilon$. By doing so, B can be searched via an iterative process until convergence. The gradients of functions $N(B)$ and $E(B)$ can be obtained as (refer to lemmas 1–2)

$$\frac{\partial N(B)}{\partial B} = 4\left[\mathrm{tr}\left(B^{\mathrm{T}}B\right) - M\right] \cdot B \tag{5}$$

and

$$\frac{\partial E(B)}{\partial B} = -\int_0^{t_f} \mathrm{e}^{A^{\mathrm{T}}t}W_B^{-1}X_{\mathrm{f}}W_B^{-1}\mathrm{e}^{At}\mathrm{d}t\ B \tag{6}$$

respectively. The two gradients point to the directions where $N(B)$ and $E(B)$ have the fastest increasing speed respectively. We also show that $\frac{\partial E(B)}{\partial B}$ is always non-zero as long as (A, B) is controllable (theorem 1), which explains the importance of imposing the normalization condition on B.

Let $\{\cdot\}$ denote a vector form of an argument matrix constructed by stacking its columns. We define a projection operator onto the tangent space of $\{\frac{\partial N(B)}{\partial B}\}$ as

$$P_{\frac{\partial N(B)}{\partial B}}(v) = \left(I - \left\{\frac{\partial N(B)}{\partial B}\right\}\left\{\frac{\partial N(B)}{\partial B}\right\}^{+}\right)(v) \tag{7}$$

where v denotes a particular vector and $^{+}$ is the Moore–Penrose pseudo-inverse of a matrix. More detailed information of the defined projection operator can be found in lemma 3.

Projected gradient method (PGM). Let $\Gamma(B)$ be an operator which normalizes B onto the sphere surface $\mathrm{tr}(B^{\mathrm{T}}B) = M + \epsilon$. Note that, to guarantee the controllability of a complex network, the minimum number of controllers is the number of driver nodes N_{D}, which can be obtained by the maximum-matching algorithm proposed in [17].

Step 1. Initiate B as a random matrix B_0 with a dimension $N \times M$ where $M \geqslant N_{\mathrm{D}}$.

Step 2. Calculate the gradients of $\frac{\partial E(B)}{\partial B}$ and $\frac{\partial N(B)}{\partial B}$, respectively. Let

$$B_{k+1} = \Gamma\left[B_k - \eta \cdot P_{\frac{\partial N(B)}{\partial B}}\left(\left\{\frac{\partial E(B)}{\partial B}\right\}\right)\right]$$

at $B = B_k$, where η is a chosen step length.

Step 3. Obtain the vector version of $\frac{\partial E(B)}{\partial B}$ and $\frac{\partial N(B)}{\partial B}$, denoted as $\left\{\frac{\partial E(B)}{\partial B}\right\}$ and $\left\{\frac{\partial N(B)}{\partial B}\right\}$ respectively, and calculate the angle θ_k between them at $B = B_k$.

Step 4. Update $k = k + 1$. If $\cos(\theta_k) + 1 > \xi$, go to Step 2; otherwise, stop. Note that ξ refers to the convergence criteria, which can be set as a sufficiently small positive value. In the next section, we will prove that $E(B_k)$ for all k is a non-increasing sequence which converges to a non-zero local minimum, yielding a solution of B denoted by B^{s}.

3. Convergence analysis

The main idea of the proposed PGM is to obtain the derivative of functions $N(B)$ and $E(B)$ with respect to matrix B respectively, and then project the descent gradient of $E(B)$ onto the sphere surface $\mathrm{tr}(B^{\mathrm{T}}B) = M + \epsilon$; the process is repeated until the iteration converges to the solution. Some preliminaries and lemmas are presented first.

3.1. Preliminaries and lemmas

Lemma 1. *For a matrix X with no special structure (elements of X are independent), we have [28] the following.*

- $\frac{\partial [X]_{kl}}{\partial [X]_{ij}} = \delta_{ki} \cdot \delta_{lj}$.
- $\frac{\partial [X^{\mathrm{T}}]_{kl}}{\partial [X]_{ij}} = \delta_{kj} \cdot \delta_{li}$.
- *For matrices A, B, X such that* $[X]_{ij} = \sum_k [A]_{ik}\, [B]_{kj}$ *(written as* $[X]_{ij} = [A]_i k \cdot [B]_k j$*),* $X = A\,B$.
- *For a function g(U) where the matrix U is a function of another matrix X, i.e.* $U = f\,(X)$, $\frac{\partial g}{\partial [X]_{ij}} = \sum_{km} \frac{\partial g}{\partial [U]_{km}} \frac{\partial [U]_{km}}{\partial [X]_{ij}}$ *(written as* $\frac{\partial g}{\partial [X]_{ij}} = \frac{\partial g}{\partial [U]_{km}} \cdot \frac{\partial [U]_{km}}{\partial [X]_{ij}}$*).*
- $\frac{\partial [\mathrm{tr}(X^{-1} X_{\mathrm{f}})]}{\partial X} = -X^{-\mathrm{T}} X_{\mathrm{f}} X^{-\mathrm{T}}$*; where* $[X]_{ij}$ *denotes the ijth element of the matrix X and* $\delta_{ij} = 1$ *iff* $i = j$*; otherwise,* $\delta_{ij} = 0$.

Lemma 2. *The gradient of the function E(B) is*

$$\frac{\partial E(B)}{\partial B} = -2\int_0^{t_{\mathrm{f}}} \mathrm{e}^{A^{\mathrm{T}}t} W_B^{-1} X_{\mathrm{f}} W_B^{-1} \mathrm{e}^{At} \mathrm{d}t\ B.$$

Proof. From lemma 1, we have

$$\frac{\partial\, \mathrm{tr}\left[\mathrm{W}_B^{-1}\mathrm{X}_{\mathrm{f}}\right]}{\partial B_{ij}} = \frac{\partial\, \mathrm{tr}\left[\mathrm{W}_B^{-1}\mathrm{X}_{\mathrm{f}}\right]}{\partial\left[W_B\right]_{kl}} \cdot \frac{\partial\left[W_B\right]_{kl}}{\partial B_{ij}}$$

$$= -\left[W_B^{-\mathrm{T}} X_{\mathrm{f}}\, W_B^{-\mathrm{T}}\right]_{kl} \cdot \frac{\partial\left[\int_0^{t_{\mathrm{f}}} \mathrm{e}^{At} BB^{\mathrm{T}} \mathrm{e}^{A^{\mathrm{T}}t} \mathrm{d}t\right]_{kl}}{\partial B_{ij}}$$

$$
\begin{aligned}
&= -\left[W_B^{-\mathrm{T}} X_{\mathrm{f}} W_B^{-\mathrm{T}}\right]_{kl} \cdot \frac{\partial\left[\int_0^{t_{\mathrm{f}}}\left[\mathrm{e}^{At}\right]_{km} \cdot [B]_{mn} \cdot \left[B^{\mathrm{T}}\right]_{nz} \cdot \left[\mathrm{e}^{A^{\mathrm{T}}t}\right]_{zl}\mathrm{d}t\right]}{\partial B_{ij}} \\
&= -\left[W_B^{-\mathrm{T}} X_{\mathrm{f}} W_B^{-\mathrm{T}}\right]_{kl} \cdot [\int_0^{t_{\mathrm{f}}}\left[\mathrm{e}^{At}\right]_{km} \cdot [\delta_{mi} \cdot \delta_{nj} \cdot \left[B^{\mathrm{T}}\right]_{nz} \\
&\quad - [B]_{mn} \cdot \delta_{nj} \cdot \delta_{zi}] \cdot \left[\mathrm{e}^{A^{\mathrm{T}}t}\right]_{zl}\mathrm{d}t] \\
&= -\left[W_B^{-\mathrm{T}} X_{\mathrm{f}} W_B^{-\mathrm{T}}\right]_{kl} \cdot \left[\int_0^{t_{\mathrm{f}}}\left[\mathrm{e}^{At}\right]_{ki} \cdot \left[B^{\mathrm{T}}\right]_{jz} \cdot \left[\mathrm{e}^{A^{\mathrm{T}}t}\right]_{zl}\mathrm{d}t\right] \\
&\quad -\left[W_B^{-\mathrm{T}} X_{\mathrm{f}} W_B^{-\mathrm{T}}\right]_{kl} \cdot \left[\int_0^{t_{\mathrm{f}}}\left[\mathrm{e}^{At}\right]_{km} \cdot [B]_{mj} \cdot \left[\mathrm{e}^{A^{\mathrm{T}}t}\right]_{il}\mathrm{d}t\right] \\
&= -2 \cdot \int_0^{t_{\mathrm{f}}} \sum_{kl}\left[\mathrm{e}^{A^{\mathrm{T}}t}\right]_{ik} \cdot \left[W_B^{-\mathrm{T}} X_{\mathrm{f}} W_B^{-\mathrm{T}}\right]_{kl} \cdot \left[\mathrm{e}^{At}\right]_{lz} \cdot B_{zj}\mathrm{d}t \\
&= -2 \cdot \left[\int_0^{t_{\mathrm{f}}} \mathrm{e}^{A^{\mathrm{T}}t} W_B^{-\mathrm{T}} X_{\mathrm{f}} W_B^{-\mathrm{T}} \mathrm{e}^{At}\mathrm{d}t\, B\right]_{ij}
\end{aligned} \tag{8}
$$

where i, j, k, l, m, n, z are indexes of elements in the matrix. Thus the lemma is proved. □

Lemma 3. *$[I - SS^+]$ is a projection operator onto a space that is orthogonal to space span$\{S\}$ [29, 30], where $^+$ is the Moore–Penrose pseudo-inverse of a matrix and span$\{S\}$ is spanned by the column vectors of S.*

Proof. For an arbitrary vector $x \in \mathrm{span}\{S\}$, x can be represented by the column vectors of S, i.e., $x = Sy$. As

$$\left[I - SS^+\right]x = \left[I - SS^+\right]Sy = Sy - Sy = 0, \tag{9}$$

the lemma is proved. □

3.2. Convergence properties

Theorem 1. *For a controllable (A, B), the gradient $\frac{\partial E(B)}{\partial B}$ can never be a zero matrix.*

Proof. Note that $X_{\mathrm{f}} = \mathrm{e}^{A^{\mathrm{T}}t_{\mathrm{f}}}\mathrm{e}^{At_{\mathrm{f}}}$ and let matrix $C = \mathrm{e}^{A^{\mathrm{T}}t}W_B^{-1}\mathrm{e}^{At_{\mathrm{f}}}$. As (A, B) is controllable, matrices $\mathrm{e}^{A^{\mathrm{T}}t}$ and W_B^{-1} are invertible. Hence we have

$$\int_0^{t_{\mathrm{f}}} \mathrm{e}^{A^{\mathrm{T}}t}W_B^{-1}X_{\mathrm{f}}W_B^{-1}\mathrm{e}^{At}\mathrm{d}t = \int_0^{t_{\mathrm{f}}} C^{\mathrm{T}}C\,\mathrm{d}t > 0$$

where a matrix $P > 0$ means that P is positive definite. On the other hand, since (A, B) is controllable, we must have $B \neq 0$. Thus, it is straightforward to prove that $\frac{\partial E(B)}{\partial B} = -2\int_0^{t_{\mathrm{f}}} C^{\mathrm{T}}C\,\mathrm{d}t \cdot B$ can never be a zero matrix as $\int_0^{t_{\mathrm{f}}} C^{\mathrm{T}}C\,\mathrm{d}t$ is positive definite.

□

Theorem 2. *Let $0 \leqslant \theta \leqslant 180^0$ be the angle between $\left\{\frac{\partial E(B)}{\partial B}\right\}$ and $\left\{\frac{\partial N(B)}{\partial B}\right\}$. Then*

$$\begin{cases} \theta < 90^0 & \text{if } \operatorname{tr}\left(B^{\mathrm{T}}B\right) < M, \\ \theta \text{ is indefinite} & \text{if } \operatorname{tr}\left(B^{\mathrm{T}}B\right) = M, \\ \theta > 90^0 & \text{if } \operatorname{tr}\left(B^{\mathrm{T}}B\right) > M. \end{cases} \tag{10}$$

Proof. Let $\omega = \mathrm{tr}(B^{\mathrm{T}}B) - M$, and define

$$L = \int_0^{t_{\mathrm{f}}} \mathrm{e}^{A^{\mathrm{T}}t}W_B^{-1}X_{\mathrm{f}}W_B^{-1}\mathrm{e}^{At}\mathrm{d}t. \tag{11}$$

From the proof of theorem 1, we know that L is a positive definite matrix, i.e., $L > 0$. From lemma 2 and theorem 1, we have

$$\frac{\partial E(B)}{\partial B}^{\mathrm{T}} \frac{\partial N(B)}{\partial B} = -8 \cdot \omega \cdot B^{\mathrm{T}} L B \tag{12}$$

and

$$\left\{ \frac{\partial E(B)}{\partial B} \right\}^{\mathrm{T}} \left\{ \frac{\partial N(B)}{\partial B} \right\} = -8 \cdot \omega \cdot \mathrm{tr}\left(B^{\mathrm{T}} L B \right). \tag{13}$$

Therefore the sign of $\left\{ \frac{\partial E(B)}{\partial B} \right\}^{\mathrm{T}} \left\{ \frac{\partial N(B)}{\partial B} \right\}$ is dependent on the sign of ω and this theorem holds. □

Theorem 3. *For the proposed PGM, we have that $E(B_k)$ is a non-increasing sequence which converges to a non-zero local minimum point $E^* > 0$.*

Proof. Based on figure 1, $\frac{\partial N(B)}{\partial B} |_{B=B_k}$ is the gradient of $N(B)$ at $B = B_k$, and it is a normal vector to the sphere surface. The operator $\left(I - \left\{ \frac{\partial N(B)}{\partial B} \right\} \left\{ \frac{\partial N(B)}{\partial B} \right\}^{+} \right)$ projects a vector onto the tangent plane denoted as $\frac{\partial N(B)}{\partial B}^{\perp}$. By projecting $-\frac{\partial E(B)}{\partial B} |_{B=B_k}$, which is the gradient descent direction of $E(B)$ at B_k, onto $\frac{\partial N(B)}{\partial B}^{\perp}$, we obtain

$$\vec{a}_k = \eta \cdot \left(I - \left\{ \frac{\partial N(B)}{\partial B} \right\} \left\{ \frac{\partial N(B)}{\partial B} \right\}^{+} \right) \left(-\frac{\partial E(B)}{\partial B} \right) \tag{14}$$

at the point $B = B_k$, where η is the iteration step length. Let

$$\vec{c}_k = B_k + \vec{a}_k \tag{15}$$

and apply the operator $\Gamma(.)$ to normalize $\vec{c}_k$ onto the sphere $\mathrm{tr}(B^{\mathrm{T}}B) = M + \epsilon$. We then have

$$B_{k+1} = \Gamma\left(\vec{c}_k \right) = \Gamma\left(B_k + \vec{a}_k \right). \tag{16}$$

Let $\vec{b}_k = B_{k+1} - B_k$ and $\tilde{\theta}^k$ be the angle between $\left\{ \frac{\partial \vec{E}(B)}{\partial B} \right\}$ and $\left\{ \vec{b}_k \right\}$. When $\eta \to 0$, it can be obtained that $\tilde{\theta}^k \to \theta^k$ and $0 \leqslant \theta^k \leqslant 90^0$ and hence $0 \leqslant \cos(\theta^k) \leqslant 1$. Thus we have

$$\begin{aligned} E\left(B_{k+1}\right) - E\left(B_k\right) &= -\eta \cdot \left\{ \frac{\partial E(B)}{\partial B} \right\}^{\mathrm{T}} \left\{ b_k \right\} |_{B=B_k} \\ &= -\eta \cdot \| \frac{\partial E(B)}{\partial B}^{\mathrm{T}} \| \cdot \| b_k \| \cdot \cos\left(\theta^k \right) |_{B=B_k} \leqslant 0. \end{aligned} \tag{17}$$

As $E(B_k)$ is a non-decreasing sequence and $E(B_k) > 0$, it converges to a non-zero local minimum point $E^* > 0$. The iterations end at $\|\vec{b}_k\| = 0$, which implies $\vec{a}_k = 0$, i.e., $\left\{ -\frac{\partial E(B)}{\partial B} \right\}$ concurs with the direction of $\left\{ \frac{\partial N(B)}{\partial B} \right\}$. Finally we have that $\cos(\theta_k)$, the cosine of the angle between $\left\{ \frac{\partial E(B)}{\partial B} \right\}$ and $\left\{ \frac{\partial N(B)}{\partial B} \right\}$, converges to *cos* $(\theta_k) = -1$. □

4. Results

We successfully apply the proposed PGM to find the minimum-cost control solutions for 1000-node Erdös–Rényi (ER) [31, 32] and scale-free (SF) networks [33, 34] and a list of real-life networks. The algorithm can handle networks at these and larsizes efficiently, though numerically solving the minimum-cost control problem for extra-large networks with hundreds of thousands or millions of nodes remains as a challenge (because of the need to take the inverses of huge matrices in the algorithm). Without loss of generality, we set $\epsilon = 1$ in all the experiments. In our numerical experiments, we first give a random initial B_0 on the normalized sphere surface for a fixed $M \geqslant N_{\mathrm{D}}$ and then apply PGM to find the solution. Note that, to avoid the *numerical controllability transition* [26], which means that, when a small number of key nodes is barely enough to ensure controllability, failure of finding a numerical solution may not be overcome by merely increasing numerical precision, in all our simulations we set the number of output controllers to be large enough.

The comparison results between the energy costs for the random initial B_0 and the B^{s} obtained by PGM for the ER and SF networks (with an average nodal degree $\mu = 4$) are summarized in table 1. It can be seen that (i) the control cost decreases as the number (M) of inputs increases and PGM steadily lowers the control cost by

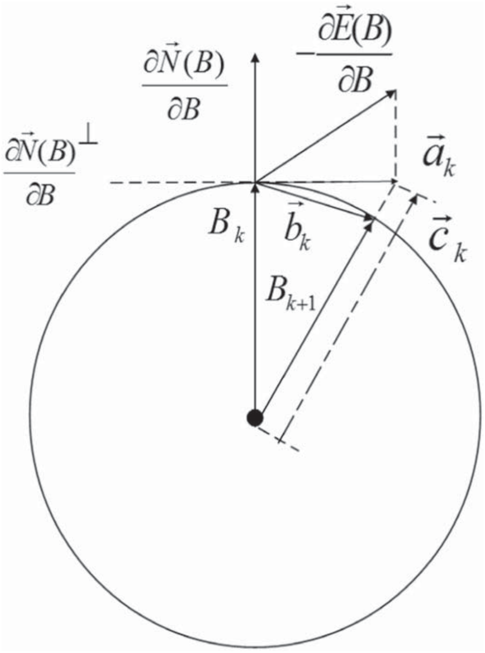
$\frac{\partial \vec{N}(B)}{\partial B}$
$-\frac{\partial \vec{E}(B)}{\partial B}$
$\frac{\partial \vec{N}(B)}{\partial B}^{\perp}$
$\vec{a}_k$
B_k
$\vec{b}_k$
$\vec{c}_k$
B_{k+1}

Figure 1. Optimization on a high dimensional sphere surface. $\frac{\partial \vec{E}(B)}{\partial B}$, $\frac{\partial \vec{N}(B)}{\partial B}$ and $\frac{\partial \vec{N}(B)}{\partial B}^{\perp}$ denote the unit vectors of $\frac{\partial E(B)}{\partial B}$, $\frac{\partial N(B)}{\partial B}$ and $\frac{\partial N(B)}{\partial B}^{\perp}$, respectively.

Table 1. Energy costs in the ER and SF networks. $\mu = 4$.

$N = 1000$	$M = 260$	$M = 290$	$M = 320$	$M = 350$
ER (initial)	1.8×10^8	3.2×10^7	7.1×10^6	1.8×10^6
PGM $t_f = 1$	7.4×10^5	2.4×10^5	7.1×10^4	2.9×10^4
PGM $t_f = 2$	3.2×10^4	2.0×10^4	7.9×10^3	5.1×10^3
SF (initial)	5.1×10^8	8.4×10^7	1.6×10^7	3.5×10^6
PGM $t_f = 1$	1.2×10^6	4.2×10^5	1.0×10^5	4.3×10^4
PGM $t_f = 2$	7.7×10^4	7.5×10^4	2.0×10^4	1.0×10^4

about two to three orders of magnitude compared to that of a randomly given B_0; (ii) the control cost decreases significantly as t_f increases—in other words, when a longer time is allowed to achieve the control objective, the control cost can be significantly reduced; and (iii) having more controllers helps drastically lower the control cost. It is, however, well known that installing more controllers may induce other costs. Being able to calculate the minimum control costs with different numbers of controllers helps to find the most cost-effective solutions for applications.

Note that for the non-convex optimization problem (4), the proposed PGM method, like any optimization method with a reasonably low complexity, can only guarantee convergence to a local minimum. An interesting observation we made in our extensive implementations of PGM starting with different initial B_0 matrices, however, is that the solutions of different rounds of implementations typically lead to nearly the same control cost. Whether this means that the solution found by a single round implementation of PGM is steadily close to the global optimum requires further careful studies. Unless otherwise specified, hereafter we present the results of an implementation of PGM with a randomly given B_0 matrix.

Table 2 demonstrates the satisfactory performance of PGM in some real-life networks, again by comparing the control costs corresponding to B_0 and B^s respectively. We observe that, with a small value of $t_f = 1$, PGM lowers the control cost by up to three to five orders of magnitude in these networks compared to that of their random counterparts.

Table 3 presents in more detail the impacts of t_f and M on the energy cost in real-life networks.

Table 2. Energy costs in real-life networks. $t_f = 1$.

Dataset	Networks	N	N_D	M	Initial	PGM
Electronic circuit [35]	Circuit-s838	512	119	150	8.5×10^8	9.2×10^5
	Circuit-s420	252	59	70	1.3×10^9	1.1×10^6
	Circuit-s208	122	29	35	6.5×10^8	3.2×10^5
Food web [36–38]	Michigan	39	13	15	3.5×10^5	990
	Phode	25	8	9	9.3×10^5	2.6×10^3
	Maspalomas	30	9	10	2.5×10^9	1.0×10^4
Social influence [39]	Phys-discuss-rev	231	85	100	2.5×10^8	1.2×10^4
Social [40]	cons-freq-rev	46	2	10	189.0	7.6
	Phys-friend-rev	228	52	60	2.0×10^{10}	7.9×10^6

It can be seen that, as that in the ER and SF networks, increasing t_f and M helps significantly lower the control cost. Also, in most cases, PGM lowers the control cost by a few orders of magnitude compared to that of the initial random connections.

It is interesting to note that similar observations have been made in [21, 22] that increasing the number of controllers helps significantly lower the control cost. Our results show that the conclusion holds in the minimum-cost optimal solutions. As an example, figure 2 illustrates in detail the relation between the number of control nodes and the corresponding minimum control cost in a real-life network.

Note that both B_0 and B^s require many nodes each to be connected to each of the M controllers, which may not yield a practical solution for large-size networks. A practical and interesting problem therefore is to find the minimum-cost control solution under the constraint that only a small set of nodes can be directly connected to external controllers. We term this small node set a *key node set* and consider the case where each key node can be connected to only a unique controller to achieve the lowest connection complexity.

For this case, we propose a very simple approach to derive the key node set from B^s. Note that the absolute value of link weight $|B_{ij}^s|$ reflects the importance of the node i for the jth controller. Define an *importance index* vector

$$r = \frac{\left[r_1 \ \dots \ r_i \ \dots \ r_N\right]}{\max\left(r_1,\ \dots,\ r_i,\ \dots,\ r_N\right)} \tag{18}$$

where $r_i = \sum_j |B_{ij}^s|$ for $i = 1, \dots, N$. Clearly, we have $\max\{r\} = 1$. The importance index of node i evaluates the relative importance of this node in achieving the minimum-cost control objective. We form the *key node set* as the first M nodes with the largest importance index values. The corresponding connection matrix, denoted as B^*, can be easily constructed: if a node i is the kth node of the key node set, we set $B_{i,k}^* = 1$; otherwise, $B_{i,k}^* = 0$. We term this simple approach the PGM extension (PGME).

Table 4 shows that, while reducing the number of connections between network nodes and external controllers from a typical value of $M \times N$ of a calculated B^s to only M of its corresponding B^*, the control cost is only marginally increased. Note that, while B^s lowers the control cost typically by two to three orders of magnitude compared to that of a random B_0, B^* drastically lowers the control cost even more compared to that of the random one-to-one connections between controllers and network nodes. Specifically, in order to generate a random solution that ensures network controllability, we first apply the maximum matching method [17] to find one set of driver nodes, which could be any one among the multiple maximum matching solutions for the network. Then we randomly choose $M - N_D$ additional nodes in the network to construct the control node set. We term this simple method the *random allocation method* (RAM).

Example cases of PGME results versus the average of 100-times implementations of RAM are plotted in figure 3. It is observed that PGME, compared to RAM, lowers the control cost by up to eight to nine orders of magnitude in both synthetic and real-life networks. Such a conclusion basically holds in all the other networks we have tested. In the extreme case where $M = N$, every node becomes a key node and consequently PGME and RAM become equivalent.

A closer examination of what kinds of node are selected as key nodes helps reveal some useful insights. To start, we consider two elementary topologies which widely exist in various complex networks: a stem and a circle. Examples of these two elementary topologies, each of which contain six nodes, are illustrated in figure 4. We see that the key nodes tend to divide the elementary stem and circle equally. A few other simple topologies have been tested and the conclusion always holds. This observation may be of importance, as it may lead to heuristic algorithm design for the minimum-cost control of extra-large complex networks.

Table 3. The effects of t_f and M on energy cost in real-life networks.

Dataset	Networks	t_f	N	M_1	Initial	PGM	M_2	Initial	PGM
Electronic circuit	Circuit-s838	1	512	150	8.5×10^8	9.2×10^5	190	7.9×10^6	5.0×10^4
		2	512	150	8.3×10^6	2.2×10^4	190	3.2×105	3.2×10^3
	Circuit-s420	1	252	70	1.3×10^9	1.1×10^6	92	5.0×10^6	2.6×10^4
		2	252	70	1.8×10^7	1.8×10^4	92	1.8×10^5	1.6×10^3
	Circuit-s208	1	122	35	6.5×10^8	3.2×10^5	45	2.7×10^6	9.6×10^3
		2	122	35	3.0×10^6	5.4×10^3	45	7.2×10^4	6.0×10^2
Food web	Michigan	1	39	15	3.5×10^5	1.0×10^3	17	3.3×10^5	7.3×10^2
		2	39	15	1.4×10^4	1.4×10^2	17	1.3×10^4	91.5
	Rhode	1	25	9	9.3×10^5	2.6×10^3	10	2.3×10^5	8.6×10^2
		2	25	9	3.8×10^5	1.8×10^2	10	9.6×10^4	86.6
	Maspalomas	1	30	10	2.5×10^9	1.0×10^4	12	1.1×10^6	9.6×10^2
		2	30	10	7.1×10^4	4.3×10^2	12	4.2×10^4	91.7
Social influence	Phys-discuss-rev	1	231	100	2.5×10^8	1.2×10^4	107	1.0×10^7	5.8×10^3
		2	231	100	2.8×10^6	1.1×10^3	107	3.3×10^5	7.3×10^2
Social	cons-freq-rev	1	46	10	369.0	7.6	16	185.80	4.90
		2	46	10	3.91	2.38	16	2.46	1.54
	Phys-friend-rev	1	228	60	2.0×10^{10}	7.9×10^6	81	7.9×10^7	7.2×10^4
		2	228	60	9.7×10^7	9.0×10^4	81	2.0×10^6	3.3×10^3

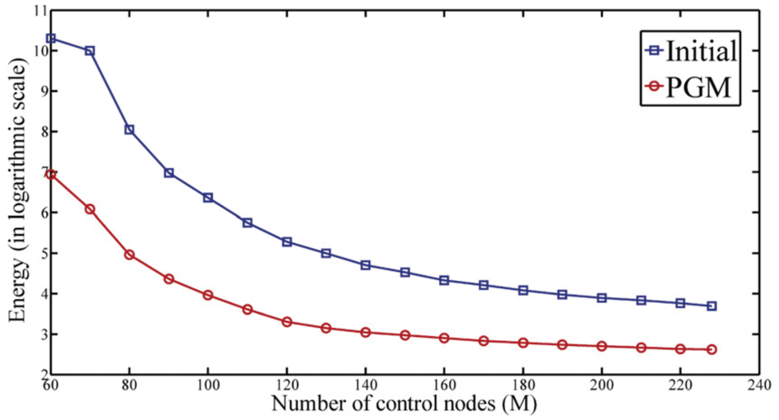

Figure 2. Number of control nodes versus the energy consumptions in a real-life (frequency-rev [40]) network.

Table 4. $E(B^s)$ and $E(B^*)$ in the ER/SF/real-life networks.

Networks	N	M	$E(B^s)$	$E(B^*)$
ER ($\mu = 4$)	1000	260	1.821×10^8	1.822×10^8
		290	7.146×10^6	7.150×10^6
SF ($\mu = 4$)	1000	260	5.173×10^8	5.173×10^8
		290	1.625×10^7	1.625×10^7
Elementary stem	100	20	5.85×10^8	6.47×10^8
		25	2.82×10^6	3.11×10^6
Elementary circle	100	20	5.85×10^8	6.47×10^8
		25	2.82×10^6	3.11×10^6

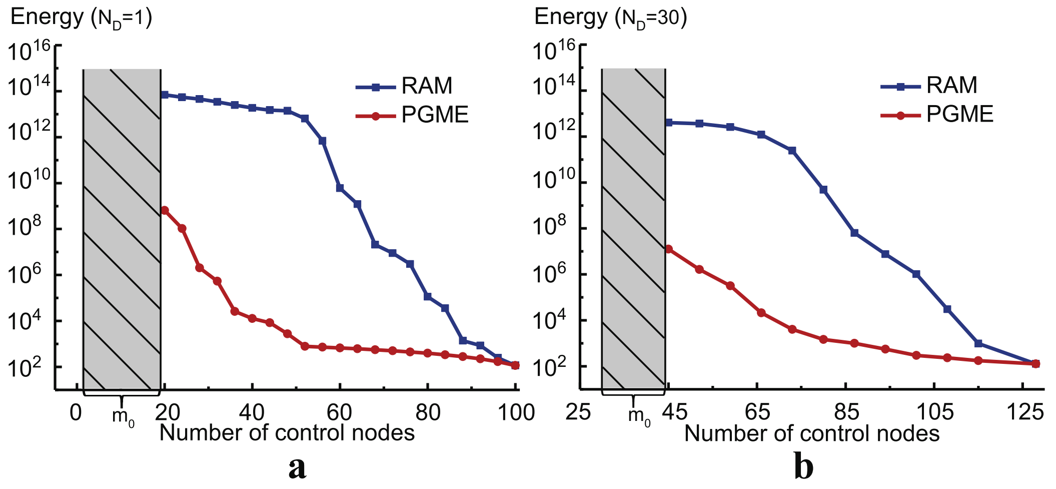

Figure 3. Comparisons between PGME and RAM in synthetic and real-life networks. The shaded part denotes the 'numerical controllability transition' area and $M = N_D + m_0$. Comparison results in (a) a synthetic network: an elementary stem with $N = 100$; and (b) a real-life electronic network (circuit-s208) with $N = 128$ [27].

Another even more important observation is that the key node set basically avoids high-degree nodes in both synthetic and real-life networks, as shown in table 5. For the ER/SF networks, since the mean degrees of both networks are around 4, the sum of the mean input and output degrees, denoted by $\mu^{\mathrm{in}+\mathrm{out}}$, will be around 8. Surprisingly, the sums of the mean input and output degrees of the key nodes identified by PGME, denoted as $\mu_{\mathrm{key}}^{\mathrm{in}+\mathrm{out}}$, are only around 5.52 and 4.70 in the two networks respectively. For the real-life networks, the same conclusion holds: the $\mu_{\mathrm{key}}^{\mathrm{in}+\mathrm{out}}$ values identified by PGME are much lower than the average degree of the top 5% of highest-degree nodes; basically no hub node is selected as the key node. This suggests that the key node set is unlikely to be the hubs in the networks. In [17], it was pointed out that the driver node set which minimizes the *number* of nodes directly connected to controllers avoids the hub nodes. It is interesting to see that the key node set which minimizes the overall control *cost* also avoids hub nodes. It is the low-degree nodes at the correct locations that determine the cost for driving a complex network to any state. Such an observation may have very important implications, as it may change our basic understanding and approaches of complex system control.

Last, it is worth mentioning that, starting from any feasible initial (A, B_0), solution of the PGM method guarantees network controllability as the network control cost remains strictly non-increasing in each iteration of the algorithm. It is however difficult to prove that the same conclusion holds for PGME. Observing in our extensive simulation experiences that PGME always generates a feasible solution with comparable performance to that of PGM, we have good reasons to expect that solution of PGME would almost always guarantee the controllability of the system, especially when we need in any case a large enough number of control nodes to avoid the numerical controllability transition area.

5. Conclusion

There are three main contributions/observations reported in this paper: (1) the proposed PGM method for the first time provides a solution to the important open problem of realizing the minimum-cost control of complex networks; (2) it is revealed that the complexity of the connections between the external controllers and network nodes can be minimized without a significant increase in control cost; and (3) it is found that the key nodes leading to the minimum control cost are basically low-degree nodes. It is exciting to know that a set of properly selected low-degree nodes, with the simplest one-to-one connections with the external controllers, can achieve a suboptimal solution in minimizing the cost for controlling complex networks; and we have a first algorithm that can efficiently find such a set of nodes. The proposed algorithm and the achieved insights will be of significant importance for various applications. In this direction, many further studies can be carried out, e.g. to develop simpler heuristic algorithms for finding the key node set in extra-large networks, or to investigate the identification of the key node set without accurate global network topology information, etc.

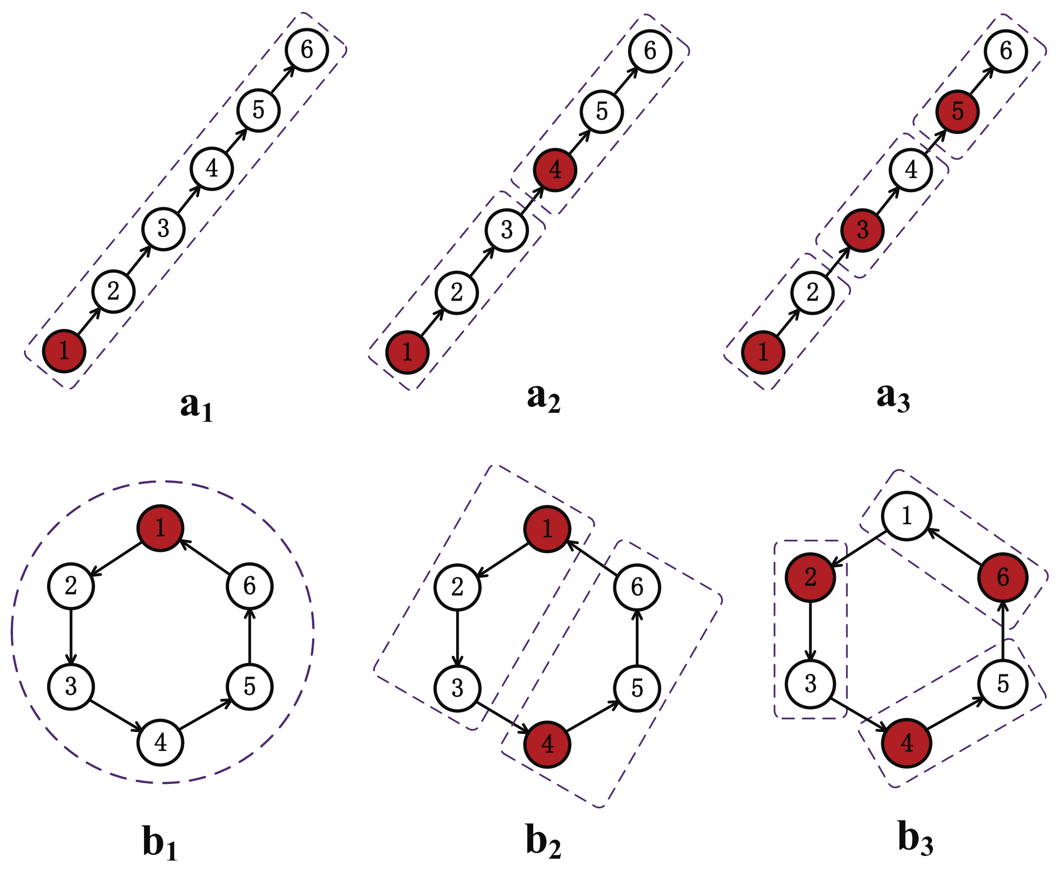

Figure 4. Key nodes in two elementary topologies. (a) An elementary stem where $N = 6$ and $M = 1, 2, 3$. (b) An elementary circle where $N = 6$ and $M = 1, 2, 3$.

Table 5. $\mu_{\text{key}}^{\text{in+out}}$ in ER/SF/real-life networks.

Networks	M	$\mu_{\text{max}}^{\text{in+out}}$	$\mu_{\text{top5\%}}^{\text{in+out}}$	$\mu_{\text{key}}^{\text{in+out}}$
ER	260	26	18.2	5.52
	290	26	18.2	5.64
	320	26	18.2	5.72
	350	26	18.2	6.02
SF	260	124	68.3	4.70
	290	124	68.3	4.78
	320	124	68.3	4.87
	350	124	68.3	5.02
Circuit-s208	35	10	7.5	5.52
Circuit-s420	70	14	7.62	5.64
Circuit-s838	35	22	7.65	5.72
Rhode	9	18	15	6.02
Michigan	15	43	36	4.70
Maspalomas	10	21	19.5	4.78
discuss-rev	100	21	11.75	4.87
friend-rev	10	13	10.42	5.02
frequency-rev	46	65	64	39

Acknowledgments

The work was partially supported by the Independent Research Plan of Tsinghua University (20141080934 and 20151080467), and the project sponsored by SRF for ROCS, SEM, and Ministry of Education, Singapore, under contracts RG28/14 and MOE2013-T2–2-006. Part of this work is an outcome of the Future Resilient Systems project at the Singapore-ETH Centre (SEC), which is funded by the National Research Foundation of Singapore (NRF) under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] Newman M, Watts D and Strogatz S 2002 *Proc. Natl Acad. Sci.* **99** (Suppl. 1) 2566–72
[2] Kramer A D I, Guillory J E and Hancock J T 2014 *Proc. Natl Acad. Sci.* **111** 8788–90
[3] Nightingale A J and Cote M 2012 *Prog. Hum. Geogr.* **36** 470–89
[4] Daron J D, Sutherland K and Jack C 2015 *Regional Environ. Change* **15** 1–2
[5] Roy S, Xue M and Das S K 2012 *IEEE Trans. Parall. Distr. Syst.* **23** 1694–707
[6] Fawzi H, Tabuada P and Diggavi S 2014 *IEEE Trans. Autom. Control* **59** 1454–67
[7] Chen Y, Lu J, Yu X and Hill D J 2013 *IEEE Circuits Devices Mag.* **13** 21–34
[8] Dibaji S M and Ishii H 2015 *Systems Control Lett.* **79** 23–9
[9] Yang P *et al* 2010 *Automatica* **46** 390–6
[10] Liu B, Dousse O, Nain P and Towsley D 2012 *IEEE Trans. Parall. Distr. Syst.* **24** 301–11
[11] Zavlanos M M *et al* 2011 *Proc. IEEE* **99** 1525–40
[12] Gouveia B D, Portugal D, Silva D C and Marques L 2014 *IEEE Trans. Autom. Sci. Eng.* **12** 410–22
[13] Lo C and Ansari N 2012 *IEEE Commun. Surv. Tutor* **14** 799–821
[14] Gao J, Xiao Y, Liu J, Liang W and Chen C L P 2012 *Future Gener. Comput. Syst.* **28** 391–404
[15] Ruths J and Ruths D 2014 *Science* **343** 1373
[16] Lin C 1974 *IEEE Trans. Autom. Control* **19** 201
[17] Liu Y, Slotine J and Barabasi A 2011 *Nature* **473** 167
[18] Yuan Z *et al* 2013 *Nat. Commun.* **4** 2447
[19] Gao J *et al* 2014 *Nat. Commun.* **5** 5415
[20] Menichetti G, DallAsta L and Bianconi G 2014 *Phys. Rev. Lett.* **113** 078701
[21] Yan G, Tsekenis G, Barzel B, Slotine J J, Liu Y Y and Barabasi A L 2015 *Nat. Phys.* **11** 779
[22] Chen Y Z, Wang L, Wang W and Lai Y C 2015 arXiv:1509.03196 [cs.SY]
[23] Yan G, Ren J, Lai Y, Lai C and Li B 2012 *Phys. Rev. Lett.* **108** 218703
[24] Rugh W J 1996 *Linear System Theory* (Upper Saddle River, NJ: Prentice-Hall)
[25] Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H and Herwig R 2009 *Systems Biology: A Textbook* (Weinheim: Wiley)
[26] Sun J and Motter A E 2013 *Phys. Rev. Lett.* **110** 208701
[27] Milo R, Itzkovitz S, Kashtan N, Levitt R and Shen-Orr S 2004 Superfamilies of evolved and designed networks *Science* **303** 1538–42
[28] Petersen J and Pedersen J 2012 *The Matrix Cookbook* Version 16 February 2012
[29] Li G, Wen C, Zheng W and Chen Y 2011 *IEEE Trans. Signal Process.* **59** 2146
[30] James M 1978 *Math. Gazette* **62** 109
[31] Erdos P and Renyi A 1960 *Magyar Tud. Akad. Mat. Kutato Int. Kozl.* **5** 17
[32] Buldyrev S V, Parshani R, Paul G, Stanley H E and Havlin S 2010 *Nature* **464** 1025–88
[33] Albert R and Barabsi A 2002 *Rev. Mod. Phys.* **74** 47
[34] Barabasi A and Bonabeau E 2003 *Sci. Am.* **288** 50–9
[35] Milo R *et al* 2004 *Science* **303** 1538
[36] Christian R and Luczkovich J 1999 *Ecol. Modelling* **117** 99
[37] Monaco M E and Ulanowicz R E 1997 *Mar. Ecol. Prog. Ser.* **161** 239
[38] Baird D, Luczkovich J and Christian R 1998 *Estuar. Coast. Shelf Sci.* **47** 329
[39] Leskovec J, Kleinberg J and Faloutsos C 2007 *ACM Trans. Knowl. Discov. Data* **1** 2
[40] Burt R S 1987 *Am. J. Sociol.* **92** 1287