Chapter 5
The Bow-Tie Model of Ownership Networks

Indeed, even some of the very simplest programs that I looked at had behavior that was as complex as anything I had ever seen.
It took me more than a decade to come to terms with this result, and to realize just how fundamental and far-reaching its consequences are.

(S. Wolfram in Wolfram 2002, p. 2)

Perhaps the most surprising feature discovered in the empirical network analysis of Chap. 4 is the emergence of a tiny, powerful, tightly-knit and self-controlled group of corporations, see Sect. 4.3.5. This core can be identified as the strongly connected component (SCC[1]) of an emerging bow-tie structure[2] in the global network of TNCs, located in the largest connected component (LCC) of the network, see Sect. 4.2.4. Collectively this core holds close to 40 % of the total control in the network, despite being comprise of only 1347 corporations. Recall that the network size is 600508. The relevance of this structure is discussed in Sect. 6.1.6 and its implications in Sect. 6.2. The emergence of a bow-tie topology in the global ownership network, has, to our knowledge, never been observed before. Recall also that in the cross-country analysis of Chap. 3 we also uncovered bow-tie structures in various national networks, see Sect. 3.3.1.

It is known that technological networks, such as the World-Wide Web (WWW) (Broder et al. 2000) and Wikipedia[3] (Capocci et al. 2006), also exhibit bow-tie topologies. In contrast to the ownership networks, their SCCs are large, comprising more than half of all the nodes.

What are the organizational principles and the driving forces behind this kind of network organization? In order to understand the mechanisms underlying the formation of different bow-tie structures, we develop a generic Modeling Framework in Sect. 5.2. It is governed by node and link addition, where the network evolution

[1] A list of acronyms can be found in front matter of this book.

[2] Recall Figs. 1.2 and 3.1.

[3] The Internet encyclopedia, http://wikipedia.org.

J. B. Glattfelder, *Decoding Complexity*, Springer Theses,
DOI: 10.1007/978-3-642-33424-5_5, © Springer-Verlag Berlin Heidelberg 2013

is determined by a preferential-attachment mechanism defined by a distribution of fitness values amongst the nodes. This fitness measure can either be determined by the network topology (e.g., degree, centrality, network control, etc.) or can be a non-topological state variable (e.g., operating revenue). The network formation is determined by the co-evolution of fitness and topology: at each time step, the distribution of fitness determines the topology of the network, which in turn impacts the distribution of fitness in the next step.

Using the Modeling Framework we can address the question of what the simplest mechanisms are that result in the emergence of bow-tie structures. In other words, what interactions are necessary at the micro-level in order to reproduce the observed macro-patterns. In detail, in Sect. 5.3, we present a specific incarnation of the framework to reproduce the empirical properties of the TNC network. We focus on the specific sizes of the bow-tie components and the degree distribution. This sheds new light on the possible interaction-mechanisms of economic agents in ownership networks.

First, in Sect. 5.1, we discuss some general network-theoretical aspects of bow-tie topologies in networks where the nodes are uncorrelated. These insights can be understood as our null hypothesis: the emergence of bow-tie structures in random networks.

5.1 Bow-Tie Components Size of Networks

What structures can be expected in generic, uncorrelated networks: can random networks exhibit bow-tie topologies? Note that the Appendices B.6.1 and B.6.2 cover some details of undirected and directed random graphs, respectively.

5.1.1 Theoretical Components Size of Directed Networks

To summarize, in directed networks, a (weakly) connected component (CC) refers to the set of nodes and vertices that are connected regardless of the direction of the links. Similarly, a strongly SCC is a subgraph in which each node is reachable from every other node by a chain of directed links.

Within any CC there could be many SCCs. By choosing a SCC, it is possible to define a bow-tie topology, with the SCC as its core. Furthermore, there often exists a large connected component next to smaller disconnected components in arbitrary networks. The LCCs largest SCC (LSCC) unambiguously defines the predominant bow-tie structure.

The size of the LCC for an undirected graph can be computed from the following heuristic argument. Let u be the fraction of nodes not in the LCC. The probability that

a node i is not in the LCC is equal to the probability that none of its neighbors belong to the LCC, which is given by u^{k_i}. Thus the following self-consistency relation holds

$$u = \sum_k \mathcal{P}(k) u^k =: \phi(u). \tag{5.1}$$

Note that the right-hand side of Eq. (5.1) defines a so-called generating function $\phi(u)$ (see (Durrett 2004) for details). The size of the LCC, denoted by $\mathcal{L}$, can be computed from the relations $\mathcal{L} = 1 - \phi(u)$ and $\phi(u) = u$, as will be demonstrated below for the directed case.

Similarly, using the generating function formalism for directed graphs (Newman et al. 2001; Dorogovtsev et al. 2001)

$$\phi(x, y) := \sum_{k_{in}, k_{out}} \mathcal{P}(k_{in}, k_{out}) x^{k_{in}} y^{k_{out}}, \tag{5.2}$$

the components size can be computed analytically as a function of the link probability $p_{tot} = z/n$, where $z = \langle k_{tot} \rangle$ in the directed case. Note that $\zeta := z/2 = z_{in} = z_{out}$. Furthermore

$$\zeta = \partial_x \, \phi(x, 1)|_{x=1} = \partial_y \, \phi(1, y)|_{y=1} \, . \tag{5.3}$$

Defining

$$\phi_1^{LCC}(x) := \frac{1}{z} \partial_x \phi(x, x), \tag{5.4}$$

$$\phi_1^{IN}(x) := \frac{1}{\zeta} \partial_y \, \phi(x, y)|_{y=1} \, , \tag{5.5}$$

$$\phi_1^{OUT}(y) := \frac{1}{\zeta} \partial_x \, \phi(x, y)|_{x=1} \, , \tag{5.6}$$

the components size obey

$$\mathcal{L} := 1 - \phi(x_c, x_c); \quad x_c = \phi_1^{LCC}(x_c), \tag{5.7}$$

$$\widehat{\mathcal{I}} := 1 - \phi(x_c, 1); \quad x_c = \phi_1^{IN}(x_c), \tag{5.8}$$

$$\widehat{\mathcal{O}} := 1 - \phi(1, y_c); \quad y_c = \phi_1^{OUT}(y_c). \tag{5.9}$$

Note that the values $\widehat{\mathcal{I}}$ and $\widehat{\mathcal{O}}$ both also contain the size of the SCC, denoted by $\mathcal{S}$, which is given by

$$\begin{aligned} \mathcal{S} &:= 1 - \phi(x_c, 1) - \phi(1, y_c) + \phi(x_c, y_c) \\ &= \widehat{\mathcal{I}} \cdot \widehat{\mathcal{O}} - \phi(x_c, 1) \cdot \phi(1, y_c) + \phi(x_c, y_c). \end{aligned} \tag{5.10}$$

This means that the sizes of the IN and OUT components is given by

$$\mathcal{I} := \widehat{\mathcal{I}} - \mathcal{S}, \tag{5.11}$$

$$\mathcal{O} := \widehat{\mathcal{O}} - \mathcal{S}. \tag{5.12}$$

Finally, for the tubes and tendrils (T&T) one finds

$$\mathcal{T} := \mathcal{L} + \mathcal{S} - \widehat{\mathcal{I}} - \widehat{\mathcal{O}} = \mathcal{L} - \mathcal{S} - \mathcal{I} - \mathcal{O}. \tag{5.13}$$

5.1.2 Components Size of Directed Random Networks

Applying this methodology to the case of a directed random graph, described in Appendix B.6.2, the generating function is

$$\begin{aligned}\phi(x, y) &= \sum_{k_{in},k_{out}} \mathcal{P}(k_{in}, k_{out}) x^{k_{in}} y^{k_{out}} = \sum_{k_{in},k_{out}} \mathcal{P}(k_{in})\mathcal{P}(k_{out}) x^{k_{in}} y^{k_{out}} \\ &= \sum_{k_{in}} \frac{(\zeta x)^{k_{in}} e^{-\zeta}}{k_{in}!} \sum_{k_{out}} \frac{(\zeta y)^{k_{out}} e^{-\zeta}}{k_{out}!} = e^{-z+\zeta(x+y)}.\end{aligned} \tag{5.14}$$

Hence

$$\phi_1^{LCC}(x) = e^{z(x-1)} = \phi(x, x), \tag{5.15}$$

$$\phi_1^{IN}(x) = e^{\zeta(x-1)}, \tag{5.16}$$

$$\phi_1^{OUT}(y) = e^{\zeta(y-1)}. \tag{5.17}$$

From Eq. (5.7) the LCC size is found to be

$$\mathcal{L} = 1 - e^{z(x_c-1)}, \tag{5.18}$$

$$x_c = e^{z(x_c-1)}, \tag{5.19}$$

and it holds that $x_c = 1 - \mathcal{L}$, which, substituted into Eq. (5.18), yields

$$\mathcal{L} = 1 - e^{-z\mathcal{L}}. \tag{5.20}$$

Correspondingly, the IN plus SCC size, respectively the OUT plus SCC size are

$$\widehat{\mathcal{I}} = 1 - e^{-\zeta\widehat{\mathcal{I}}}, \tag{5.21}$$

$$\widehat{\mathcal{O}} = 1 - e^{-\zeta\widehat{\mathcal{O}}}. \tag{5.22}$$

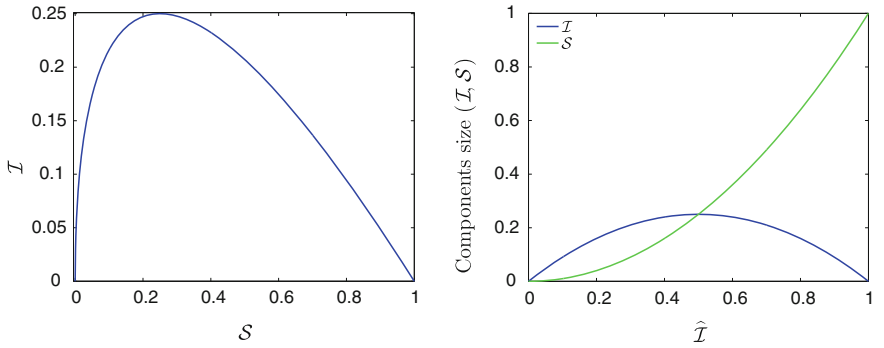

Fig. 5.1 Bow-tie components size as functions of each other: (*left*) $\mathcal{I}(\mathcal{S})$ from Eq. (5.25); (*right*) $\mathcal{I}$ and $\mathcal{S}$ as functions of $\widehat{\mathcal{I}}$ from Eqs. (5.23) and (5.26), respectively

This reveals that $\widehat{\mathcal{I}} = \widehat{\mathcal{O}}$. Equations. (5.20)–(5.22) are transcendental with no closed-form solutions. They can be numerically solved by computing the zeros of the function $f(\mathcal{C}) = 1 - e^{-z\mathcal{C}} - \mathcal{C}$, $\mathcal{C}$ being the component's size, for varying z.

Note that because Eq. (5.14) factorizes (see Appendix B.6.2), Eq. (5.10) yields

$$\mathcal{S} = \widehat{\mathcal{I}}^2 = \widehat{\mathcal{O}}^2. \tag{5.23}$$

Once the size of the SCC is determined, $\mathcal{S} = \mathcal{S}(\widehat{\mathcal{I}}, \widehat{\mathcal{O}})$ the sizes of $\mathcal{I} = \mathcal{I}(\mathcal{S})$ can be computed from Eqs. (5.11) and (5.23) as

$$\mathcal{I}^2 + 2\mathcal{I}\mathcal{S} + \mathcal{S}^2 - \mathcal{S} = 0, \tag{5.24}$$

with the positive solution

$$\mathcal{I} = -\mathcal{S} + \sqrt{\mathcal{S}}. \tag{5.25}$$

The maximum of $\mathcal{I}(\mathcal{S})$ is reached for $\mathcal{S}^* = 0.25$, and $\mathcal{I}(\mathcal{S}^*) = 0.25$. The left-side diagram in Fig. 5.1 shows a plot of Eq. (5.25). It is also straightforward to express $\mathcal{I} = \mathcal{I}(\widehat{\mathcal{I}})$, i.e.,

$$\mathcal{I} = \widehat{\mathcal{I}} - \widehat{\mathcal{I}}^2, \tag{5.26}$$

which is shown in the right-side diagram of Fig. 5.1. The two functions in Eqs. (5.23) and (5.26) are equal if $\widehat{\mathcal{I}} = 0.5$. In other words, $\mathcal{I}(0.5) = \mathcal{S}(0.5) = 0.25$. Note that identical relations hold for $\mathcal{O} = \mathcal{O}(\widehat{\mathcal{O}})$.

In Fig. 5.2 the analytical values are plotted for the various components size as a function of the average degree. In addition, empirical simulation results are shown.

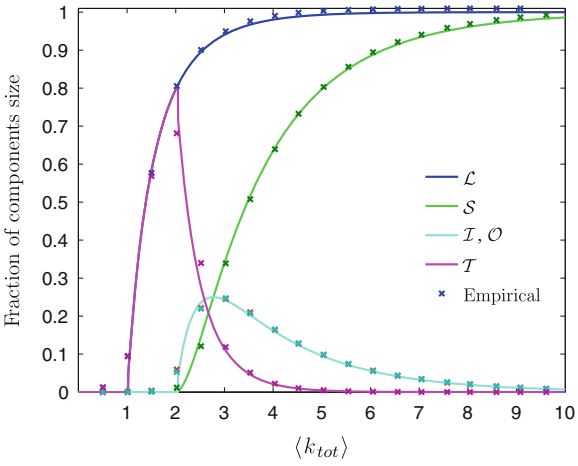

Fig. 5.2 Fraction of components size for the LCC ($\mathcal{L}$), LSCC ($\mathcal{S}$), IN ($\mathcal{I}$), OUT ($\mathcal{O}$), and T&T ($\mathcal{T}$) as a function of the total average degree $\langle k_{tot} \rangle = z = pn$; each data point denotes the average empirical values gained from 100 simulations of 1000-node networks for fixed p

From Eq. (5.21) the corresponding value of $z = 2\zeta = \langle k_{tot} \rangle$ can be computed as

$$z = -\frac{2\ln(1-\widehat{\mathcal{I}})}{\widehat{\mathcal{I}}}. \tag{5.27}$$

The maximal size of $\mathcal{I} = 0.25$ is reached at $\widehat{\mathcal{I}} = 0.5$, thus $z^*_{IN} = -4\ln(0.5) \approx 2.773$. As noted, at this point $\mathcal{I} = \mathcal{S}$.

5.1.3 Empirical TNC Components Size

We can now compare this theoretical result with the real-world global ownership network. Figure 5.3 shows the empirical components size of the TNC network for the average total degree $\langle k_{tot} \rangle \approx 3.358$.

In Table 5.1 the empirical values are compared to the theoretical ones of a directed random graph (DRG). It is apparent, that the TNC network heavily deviates from a random network.

To summarize, the following main differences between the empirically observed patterns of the TNC network and directed random graphs are observed:

- the empirical LCC size is smaller ($\mathcal{L}^{TNC} < \mathcal{L}^{DRG}$);
- only the random network has identical IN and OUT sizes, for the TNC: $\mathcal{I}^{TNC} \neq \mathcal{O}^{TNC}$;
- the TNC network's OUT size is very large ($\mathcal{O}^{TNC} > \mathcal{O}^{DRG}_{max} = 0.25$);
- the empirical IN size is very small ($\mathcal{I}^{TNC} < \mathcal{I}^{DRG}$);
- the TNC network's SCC is tiny ($\mathcal{S}^{TNC} \ll \mathcal{S}^{DRG}$);

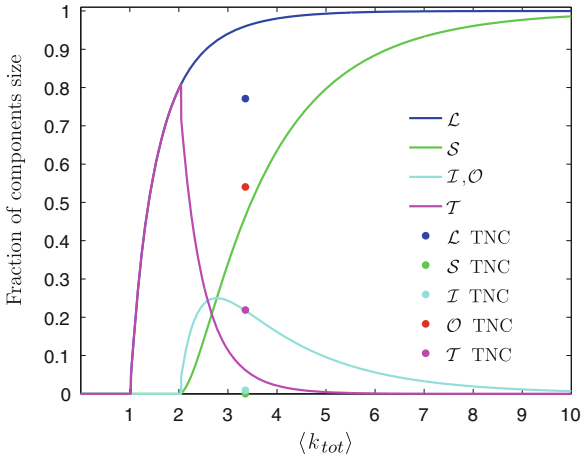

Fig. 5.3 Comparing the results of Fig. 5.2 with the components size of the TNC network shown as filled disks for the empirical value $\langle k_{tot} \rangle \approx 3.358$ and inferred $p = 5.59 \cdot 10^{-6}$

Table 5.1 Empirical and theoretical components size; the percentage values are with respect to the total network size (i.e, the LCC plus OCC)

	$\mathcal{L}$	$\widehat{\mathcal{I}}$	$\widehat{\mathcal{O}}$	$\mathcal{S}$	$\mathcal{I}$	$\mathcal{O}$	$\mathcal{T}$
TNC	0.7710	0.0095	0.5406	0.0002	0.0093	0.5404	0.2189
DRG	0.9602	0.6815	0.6815	0.4645	0.2170	0.2170	0.0616

- tiny SCC values in the DRG result in the bow-tie being mostly comprised of T&T: $\mathcal{S}^{DRG} \approx 0 \Rightarrow \mathcal{T}^{DRG} \approx \mathcal{L}^{DRG}$.

In effect, there exists no value for $\langle k_{tot}^{DRG} \rangle$ where the empirical and theoretical components size are comparable.

5.1.4 Components Size of Generalized Directed Random Networks

We now consider an extension of directed Poisson random networks that are obtained from a given *arbitrary* degree distributions for the in- and out-degree. It is customary to call such networks *generalized (directed) random networks*.

In the case of uncorrelated generalized random networks Eqs. (5.7)–(5.10) still apply and allow the components size $\mathcal{L}$, $\widehat{\mathcal{I}}$, and $\widehat{\mathcal{O}}$ to be computed. For $\mathcal{S}$, one can define

$$\mathcal{S} = \widehat{\mathcal{I}} \cdot \widehat{\mathcal{O}} + \Delta, \tag{5.28}$$

where $\Delta := -\phi(x_c, 1) \cdot \phi(1, y_c) + \phi(x_c, y_c)$ is non-zero for non-factorizable joint probability densities. Recall Appendix B.6.2.

Scale-Free Directed Networks

Scale-free networks are introduced in Appendix B.3.1. In the undirected case, their generating function is given as

$$\phi(x) = \sum_k \mathcal{P}(k)x^k = \frac{\mathrm{Li}_\alpha(xe^{-1/\kappa})}{\mathrm{Li}_\alpha(e^{-1/\kappa})}. \tag{5.29}$$

Note that due to the definition of the polylogarithm

$$\frac{d\phi(x)}{dx} = \frac{\mathrm{Li}_{\alpha-1}(xe^{-1/\kappa})}{x\mathrm{Li}_{\alpha-1}(e^{-1/\kappa})}. \tag{5.30}$$

In the directed case, it follows that the generating function takes the form

$$\phi(x, y) = \sum_{k_{in},k_{out}} \mathcal{P}(k_{in})P(k_{out})x^{k_{in}}y^{k_{out}} \tag{5.31}$$

$$= \sum_{k_{in}} \frac{k_{in}^{-\alpha}e^{-k_{in}/\kappa_{in}}}{\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})}x^{k_{in}} \sum_{k_{out}} \frac{k_{out}^{-\beta}e^{-k_{out}/\kappa_{out}}}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}y^{k_{out}} \tag{5.32}$$

$$= \frac{\mathrm{Li}_\alpha(xe^{-1/\kappa_{in}})}{\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} \frac{\mathrm{Li}_\beta(ye^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}. \tag{5.33}$$

As a first result, the average degree can be computed from Eq. (5.3)

$$\zeta = \frac{\mathrm{Li}_{\alpha-1}(e^{-1/\kappa_{in}})}{\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} = \frac{\mathrm{Li}_{\beta-1}(e^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}. \tag{5.34}$$

The minimum ζ is reached for $\alpha \to \infty$ ($\mathrm{Li}_\alpha(x) \to x$) and $\kappa \to 0$. This implies that $\zeta \geq 1$ and $z \geq 2$.

Equation (5.4) now reads

$$\phi_1^{LCC}(x) = \frac{\mathrm{Li}_{\alpha-1}(xe^{-1/\kappa_{in}})\mathrm{Li}_\beta(xe^{-1/\kappa_{out}}) + \mathrm{Li}_\alpha(xe^{-1/\kappa_{in}})\mathrm{Li}_{\beta-1}(xe^{-1/\kappa_{out}})}{zx\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}. \tag{5.35}$$

The size of the LCC is given by Eq. (5.7)

$$\mathcal{L} = 1 - \frac{\mathrm{Li}_\alpha(x_c e^{-1/\kappa_{in}})}{\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} \frac{\mathrm{Li}_\beta(x_c e^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}; \quad x_c = \phi_1^{LCC}(x_c). \tag{5.36}$$

Correspondingly, for $\widehat{\mathcal{I}}$ and $\widehat{\mathcal{O}}$

$$\phi_1^{IN}(x) = \frac{\mathrm{Li}_\alpha(xe^{-1/\kappa_{in}})}{\zeta \mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} \frac{\mathrm{Li}_{\beta-1}(e^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}, \tag{5.37}$$

$$\phi_1^{OUT}(x) = \frac{\mathrm{Li}_{\alpha-1}(e^{-1/\kappa_{in}})}{\zeta \mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} \frac{\mathrm{Li}_\beta(ye^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}, \tag{5.38}$$

and

$$\widehat{\mathcal{I}} = 1 - \frac{\mathrm{Li}_\alpha(x_c e^{-1/\kappa_{in}})}{\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})}; \quad x_c = \phi_1^{IN}(x_c), \tag{5.39}$$

$$\widehat{\mathcal{O}} = 1 - \frac{\mathrm{Li}_\beta(y_c e^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}; \quad y_c = \phi_1^{OUT}(y_c). \tag{5.40}$$

Numerical Results

In order to numerically solve the above equations, the consistency requirement of Eq. (5.34) fixes the relationship between the values α, β, κ_{in}, and κ_{out}. In other words, $\kappa_{out} = \kappa_{out}(\alpha, \beta, \kappa_{in})$ and is computed as the zeros of the function

$$K(\kappa_{out}) := \frac{\mathrm{Li}_{\alpha-1}(e^{-1/\kappa_{in}})}{\mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} - \frac{\mathrm{Li}_{\beta-1}(e^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})}, \tag{5.41}$$

i.e., $K(\hat{\kappa}_{out}) = 0$.

By defining a range of values for α, β, and κ_{in} the corresponding values of ζ and $\hat{\kappa}_{out}$ are derived. This means that the x-axis values, namely $z = 2\zeta = \langle k_{tot} \rangle$, are fixed. The numerical computation of $\mathcal{L}$, $\widehat{\mathcal{I}}$, and $\widehat{\mathcal{O}}$ is done similarly as described in Sect. 5.1.2, by finding the zeros of the following functions

$$f_{LCC}(x) := \frac{\mathrm{Li}_{\alpha-1}(xe^{-1/\kappa_{in}})\mathrm{Li}_\beta(xe^{-1/\kappa_{out}}) + \mathrm{Li}_\alpha(xe^{-1/\kappa_{in}})\mathrm{Li}_{\beta-1}(xe^{-1/\kappa_{out}})}{2\zeta x \mathrm{Li}_\alpha(e^{-1/\kappa_{in}})\mathrm{Li}_\beta(e^{-1/\kappa_{out}})} - x, \tag{5.42}$$

$$f_{IN}(x) := \frac{\mathrm{Li}_\alpha(xe^{-1/\kappa_{in}})}{\zeta \mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} \frac{\mathrm{Li}_{\beta-1}(e^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})} - x, \tag{5.43}$$

$$f_{OUT}(x) := \frac{\mathrm{Li}_{\alpha-1}(e^{-1/\kappa_{in}})}{\zeta \mathrm{Li}_\alpha(e^{-1/\kappa_{in}})} \frac{\mathrm{Li}_\beta(ye^{-1/\kappa_{out}})}{\mathrm{Li}_\beta(e^{-1/\kappa_{out}})} - x. \tag{5.44}$$

So, for instance, $f_{LCC}(x_c^{LCC}) = 0$ is a solution for $x_c^{LCC} = \phi_1^{LCC}(x_c^{LCC})$. Finally, inserting x_c^{LCC}, x_c^{IN}, and x_c^{OUT} into Eqs. (5.36), (5.39), and (5.40) for $\mathcal{L}$, $\widehat{\mathcal{I}}$, and $\widehat{\mathcal{O}}$, respectively, yields the various components size. Interestingly, they are trivial: $\mathcal{L} \equiv \widehat{\mathcal{I}} \equiv \widehat{\mathcal{O}} \equiv \mathcal{S} \equiv 1$ and $\mathcal{I} \equiv \mathcal{O} \equiv \mathcal{T} \equiv 0$. In other words, for this type of

network, only one giant SCC emerges, making-up the whole LCC, whereas all the other components are zero. This behavior is also validated by network simulations. As an example, for a scale-free network generated with $\alpha = \beta = -2.5, \kappa_{in} = \kappa_{out} = 10$, yielding $\zeta \approx 1.426$.

5.2 A Generic Network-Modeling Framework

The discussion in the last sections shows that random networks indeed can exhibit specific, regular bow-tie topologies. However, the range of variety in their components size is quite restricted. Indeed, the peculiarities of real-world complex networks with bow-tie structures, such as the TNC network, call for general mechanisms which allow the various components sizes to be tweaked arbitrarily. In order to achieve this, we move on to analyze link-formation models of networks. In other words, we move away from networks where the nodes are uncorrelated and introduce mechanisms for node-correlation. But first the existing literature on network-formation models is shortly discussed.

5.2.1 An Overview of Existing Network Models

The literature on network-formation models, yielding scale-free networks (see Appendix B.3), can perhaps best be broken down into two main strands.

Preferential-Attachment Models

The study presented in (Barabási and Albert 1999) sparked a huge wave of interest and subsequent research. For the first time it was possible to model one of the most ubiquitous properties of real-world complex networks: their scale-free nature. The proposed model was ingeniously simple and consisted of two main ingredients: growth and preferential attachment. In other words, new nodes preferably attach to existing nodes which already have many neighbors. More formally, at each time step t in the network evolution, a new node i with m links is attached to existing nodes j with the probability $\mathcal{P}(i, j) \propto k_j$, with k_j being the degree of j. Notice that this preferential-attachment model was originally devised as an undirected network model.

A multitude of generalizations and extensions to this model have been proposed. To name a few:

- adding an initial attractiveness $A > 0$ to nodes, i.e., $\mathcal{P}(i, j) \propto k_j + A$, meaning that also isolated nodes (with $k_j = 0$) can get attached to (Dorogovtsev et al. 2000;
- letting old links disappear (link decay models) (Dorogovtsev and Mendes 2000) or allowing the nodes to age (Dorogovtsev and Mendes 2000);

- adding either a new node or a new link between existing nodes at each step in the network growth (Krapivsky et al. 2001);
- a preferential-attachment mechanism with a generic non-linear attachment probability (Krapivsky and Redner 2001).

We collectively refer to this whole class of models as generalized preferential-attachment models (GPAM).

Fitness Model

On the other hand, so-called *fitness models* (Caldarelli et al. 2002; Servedio et al. 2004) build on the idea that the nodes themselves can have an intrinsic degree of freedom, called fitness. This is in strong analogy to the level-3 idea of assigning non-topological state variables to nodes, see Sect. 1.1.1. In detail, each node i has a fitness value x_i. The link-formation probability is described by an analytical function f of the fitness values, i.e., $\mathcal{P}(i, j) = f(x_i, x_j)$, where the link is formed if $f \geq \vartheta$, for some threshold ϑ. In essence, the probability distribution density of fitness $\rho(x_i)$ and the functional form of f fully determine the model. The result of these mechanisms is also a scale-free network.

In (Servedio et al. 2004) analytical methods are introduced, allowing the linking function f to be derived from the choice of the probability distribution ρ, and vice-versa. As an example, assuming $\rho(x)$ to be log-normal results in the following link-formation probability, assuming $f(x, y) = g(x)g(y)$:

$$g(x) = \sqrt{\frac{\alpha+2}{\alpha+1}\frac{\gamma^{\alpha+1}-\beta^{\alpha+1}}{\gamma^{\alpha+2}-\beta^{\alpha+2}}}\left(\beta^{\alpha+1} + (\gamma^{\alpha+1} - \beta^{\alpha+1})\mathcal{R}(x)\right)^{\frac{1}{\alpha+1}}, \qquad (5.45)$$

where $\mathcal{R}(x) = \int_0^x \rho(z)dz$. The parameter α gives the exponent of the scaling-law degree distribution $\mathcal{P}(k) \propto k^{-\alpha}$. The degree is also a function of the fitness distribution and the linking function: $k(x) = N\int_0^\infty f(x, y)\rho(z)dz$. The constants β and γ are fixed by normalization conditions ($\int_0^\infty \rho(z)dz = 1$, $g(\infty) = 1$) and are given by $\lim_{x\to 0} k(x) = \beta N$ and $\lim_{x\to\infty} k(x) = \gamma N$. N is the number of nodes in the network.

Originally, the fitness model was devised as an undirected and static model. However, it can be extended to directed networks and incorporate network growth.

Various Other Models

As can be expected, the conceivable possibilities of network-evolution models is vast. A short selection is:

- allowing nodes to belong to different types[4] (Söderberg 2002);

[4] This can be seen as a simple implementation of the idea of node fitness.

- models with specific node correlation (Vázquez et al. 2003; Boguñá and Pastor-Satorras 2003);
- weight-driven network dynamics (Barrat et al. 2004a,b);
- coupling the topology with a dynamical processes in the network (Garlaschelli et al. 2007).

5.2.2 The Bow-Tie Modeling Framework

From the previous sections it has become apparent that there is a huge variety in the proposed models of network formation. The only unifying theme to be discerned is that the link-attachment mechanism is very often driven by the degree of the nodes. In particular, models aiming at reproducing empirical networks are mostly highly tailored to the specificities and particularities of the considered real-world network. For instance, the many models of the Internet's topology show a very high degree of specialization (Zhou and Mondragón 2004; Siganos et al. 2006; Carmi et al. 2007). This is perhaps not very surprising and indicative of the fact, that there are no universally valid network models. In other words, the nitty-gritty details of each model play a crucial role in the nature of the resulting network being formed.

In focussing on the task at hand, namely devising a flexible network model that is able to reproduce varying bow-tie components size, it is an interesting observation that the existing literature on modeling empirical scale-free networks with bow-tie topologies is quite sparse. (Tadić 2001; Giammatteo et al. 2010). As mentioned, both the WWW and Wikipedia are characterized by large SCCs. In the WWW, the SCC makes up 60–70 % of the nodes, for the English-language Wikipedia, it is about 80 %.

In this thesis we present an additional strand of real-world complex networks displaying bow-tie topologies, coming from economics (Glattfelder and Battiston 2009; Vitali et al. 2011). These ownership networks are characterized by small SCCs and skewed IN and OUT sizes. In detail, the TNC network described in Chap. 4 has an interesting feature regarding the distribution of the degree in the bow tie: the nodes in the IN have high k^{out}, the nodes in the OUT high k^{in} and the SCC-nodes large k^{in} and k^{out}. This feature hints at the fact that a pure degree-driven network model is probably too rudimentary to reproduce the wealth of observed patterns.

Inspired by the approach taken in (Giammatteo et al. 2010), we propose a generic modeling framework, combining the simple elements of

1. node addition, i.e., network growth;
2. link formation between existing nodes;
3. fitness-dependent preferential attachment.

This setup can be understood as the melding of link/node growth models with fitness models, mentioned in Sect. 5.2.1. By taking a centrality measure as the nodes fitness means that now the topology of the network is driving the distribution of

Algorithm 2 ModelingFramework(P_n, P_d, m, $\mathcal{F}$, N)

Require: Initialize seed network
1: ComputeFitness($\mathcal{F}$)
2: **while** $number\_of\_nodes < N$ **do**
3: BuildNetwork(P_n, P_d, m)
4: ComputeFitness($\mathcal{F}$)
5: **end while**

fitness, which, in turn, shapes the topology in the next step of the network formation. In essence, the topology and fitness distribution co-evolve.

The Model Recipe

Algorithm 2 describes the top-level routine of the ModelingFramework(), depending on the probability of adding nodes P_n, the link-direction probability P_d, the number of links each new node brings to the network m, the final network size N and the chosen measure of fitness $\mathcal{F}$, discussed below.

At each step, the subroutine ComputeFitness($\mathcal{F}$) calculates the values of all the nodes associated with the chose fitness. The core subroutine BuildNetwork(P_n, P_d, m) contains all the modeling intelligence, given in Algorithm 3. In steps 3–16 the new node is attached to the existing network. Steps 18–21 describe the addition of a new link between existing nodes. The subroutines AddNewNode() grows the network by one node and LinkNodes(i, j) adds a link from nodes i to j. Step 19 removes the source node's neighbors from the list of possible destination nodes, as they are already linked to it and hence shouldn't be considered for adding a new link.

The subroutine SelectNodeAccordingToFitness($list\_of\_nodes$) draws a node from $list\_of\_nodes$ according to the distribution of fitness. Following (Ross 2006), this is best simulated on a computer with the pseudo-code implementation:

```
ctot = sum(ag.c);
J=1;
F = ag.c(J);
e = rand * ctot;
while(F<e)
    J = J+1;
    F = F + ag.c(J);
end
```

Where `ag` is the data structure representing the nodes (agents) properties, such as their centrality given by the vector `ag.c`, `rand` draws a uniformly distributed pseudo-random number and `sum(v)` returns the sum of the components of the vector `v`.

Algorithm 3 BuildNetwork(P_n, P_d, m)

```
1: p_n ← DrawRandomVariable()
2: if p_n ≤ P_n then
3:     new ← AddNewNode()
4:     p_d ← DrawRandomVariable()
5:     list_of_nodes ← all_nodes_in_network
6:     number_of_new_links ← 0
7:     while number_of_new_links < m do
8:         old ← SelectNodeAccordingToFitness(list_of_nodes)
9:         if p_d < P_d then
10:            LinkNodes(new, old)
11:        else
12:            LinkNodes(old, new)
13:        end if
14:        list_of_nodes ← list_of_nodes \ {old}
15:        number_of_new_links ← number_of_new_links + 1
16:    end while
17: else
18:    from ← SelectNodeAccordingToFitness(list_of_nodes)
19:    ignore ← {from} ∪ GetNeighborNodes(from)
20:    to ← SelectNodeAccordingToFitness(list_of_nodes \ {ignore})
21:    LinkNodes(from, to)
22: end if
```

This is a very general set of rules to concoct networks. The crucial ingredient is the idea, that the fitness $\mathcal{F}$ of the nodes can be any preferred measure, for instance

- k^{in} or k^{out} (for k^{in} this resembles the model in Krapivsky et al. 2001);
- the Pagerank centrality (described in Appendix B.7), yielding a simpler version[5] of the model in (Giammatteo et al. 2010);
- any of the new measures of centrality introduced in Chap. 2: $\tilde{\nu}^{\text{int}}$ and ν^{net}, or $\bar{\nu}^{\text{net}}$ and $\bar{\nu}^{\text{net}}$ (see Table 2.2).

It should be highlighted that this Modeling Framework thus reflects a true 3-level network approach.

The choices we made in defining the network evolution algorithm are kept as minimal and generic as possible. Future work could aim at tracking the impact of variations in the rules of Algorithm 21. For instance, the link addition description of steps 18–21 could be varied in many ways. The direction could also be probabilistic, introducing another modeling parameter. Or the chosen centrality could be varied for different parts of the algorithm. For instance, selecting nodes according to their in-degree if they are to be the destination of new links, or to their out-degree if they are source nodes.

[5] There the authors have further mechanisms in the link-formation process: with a certain probability, a new link between existing nodes is complemented by an opposing link, next to having a uniform distribution for the number of new links m attached to the new nodes being added.

5.2.3 Applying the Framework

To summarize, the Modeling Framework we propose comprises two dynamic elements: new nodes or new links are added in the network evolution. The recipe for choosing existing nodes to attach to is given by the node's importance, reflected in the chosen fitness measure $\mathcal{F}$.

The model is determined by four parameters:

1. the resulting size of the network N;
2. the number of links attached to the new nodes m;
3. the node/link addition probability P_n;
4. the link-direction probability P_d.

An additional choice is the shape of the initial seeding network.

$\mathcal{F}$ as Degree

The crucial parameters having the most impact on the resulting evolution of the network are P_n and P_d. To understand their meaning, we analyze a simple directed variant of the original preferential-attachment model, where the centrality is given by the total degree of the nodes plus a possible constant. In other words, the preferential-attachment probability to chose a node i is

$$\mathcal{P}(i) \propto k_i^{tot} + A = k_i^{in} + k_i^{out} + A. \tag{5.46}$$

Note that the value of A impacts the slope of the emerging scaling-law distribution of the degrees (Dorogovtsev et al. 2000; Dorogovtsev and Mendes 2003).

The simple choice $(P_n, P_d) = (1, 1)$ yields a network evolution which is only driven by node-growth, where all the new nodes point to the existing nodes in the network. Note that this results in $k^{out} \equiv m$. The behavior is equivalent for $(P_n, P_d) = (1, 0)$, switching the direction ($k^{in} \equiv m$). Figure 5.4 shows the degree distributions resulting from different values of P_d for this kind of GPAM, employing Eq. (5.46) with $A = 0$ and $P_n = 1$. When setting P_d to zero or unity, the emerging scaling-law degree probability density is, as expected from the undirected case, described by an exponent $\alpha' = \alpha - 1 \approx 3$. In essence, this corresponds to the simple generalization to directed networks of the original model proposed in (Barabási and Albert 1999) which did not consider bow-tie structures.

Observe, that as long as P_n is fixed to unity, no bow-tie structures can emerge in the whole range of $P_d \in [0, 1]$. This is by construction, then as long as the initial network contains no SCC, no loops can form in the network evolution if every new node has the direction of the m links all aligned in parallel. In contrast, if every new node would be allowed to have the direction of the m links be probabilistic, i.e., if step 4 in Algorithm 3 is moved into the while loop, we then do see the emergence of

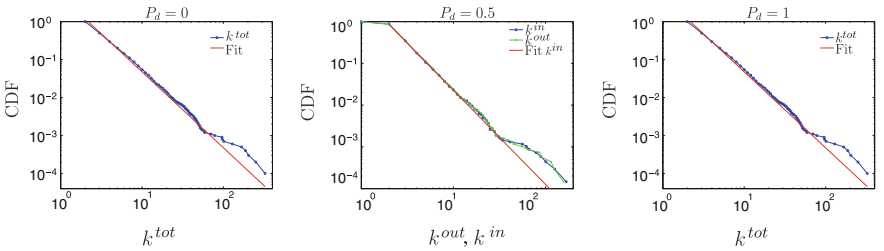

Fig. 5.4 Degree distributions of the modeling framework with the general preferential attachment probability given in Eq. (5.46), i.e., the GPAM, for $P_n = 1$, $A = 0$, $N = 10000$, $m = 2$ and by varying P_d: (*left*) $P_d = 0$, k^{tot} is depicted (note, $k^{in} \equiv m$) and a scaling law with exponent $\alpha = -2$; (*middle*) $P_d = 0.5$, k^{in} and k^{out} are show and a scaling law with exponent $\alpha = -2.25$; (*right*) $P_d = 1$, k^{tot} is shown (note, $k^{out} \equiv m$) and a scaling law with exponent $\alpha = -2$

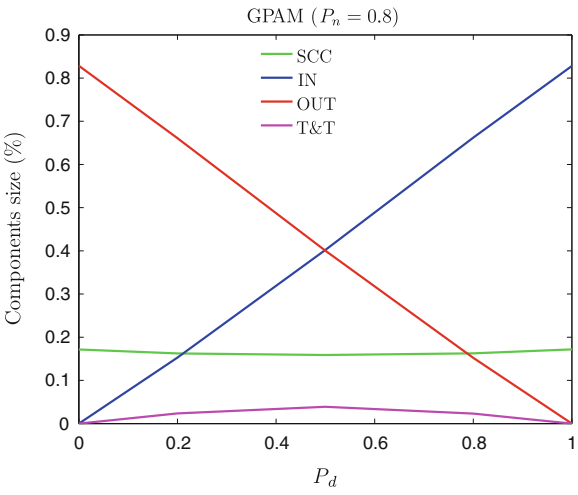

Fig. 5.5 Bow-tie components size as functions of the link direction P_d for the Modeling Framework with $P_n = 0.8$, $A = 0$ and $m = 2$ using the GPAM described in Fig. 5.4; the values are averages over ten network realizations each containing 10000 nodes

bow-tie topologies for $P_n = 1$. However, as the resulting SCC sizes are very large we don't want to investigate this variant further.

In the proposed Modeling Framework we see the first emergence of a bow-tie topology if $P_n < 1$. This is due to the fact that now new links are added to existing high-fitness nodes, which increases the probability of the emergence of an SCC. Note that the choice $P_n = 0$ is meaningless, as the network size is fixed to the size of the initial network and there is no growth. Figure 5.5 shows the quantitative dependence of the bow-tie components size as a function of P_d for $P_n = 0.8$ for the GPAM employing Eq. (5.46) and $A = 0$. A few things are interesting. The SCC has an approximately constant size, and the IN and OUT sizes have a close-to-linear behavior. There are also no sizable T&T. By setting P_n to a lower value, the SCC size increases at the cost of the other components size.

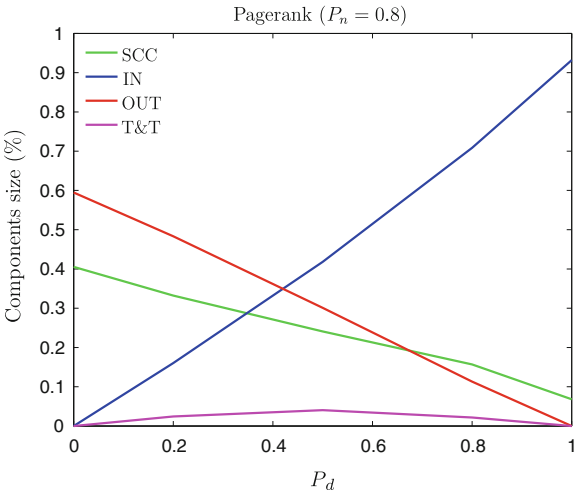

Fig. 5.6 Bow-tie components size as functions of the link direction P_d for the Modeling Framework with $P_n = 0.8$, $m = 2$ and Pagerank as generalized fitness given in Eq. (5.47); the values are averages over ten network realizations each containing 10000 nodes

These results immediately suggest that this incarnation of the Modeling Framework is already a valid candidate the reproduce the simple bow-tie sizes of the WWW and Wikipedia. The approximate size ranges of the SCC, IN and OUT are, respectively, 60–70 %, 15–20 % and 15–20 % for the WWW (Broder et al. 2000; Donato et al. 2008) and 67–90 %, 5–12 % and 4–16 % for Wikipedia (Capocci et al. 2006). Especially as the parameter A in Eq. (5.46) allows the fine-tuning of the scaling-law exponent of the degree distributions. It is encouraging that our Modeling Framework is indeed generic enough to be applicable to the modeling of two different real-world complex networks.

$\mathcal{F}$ as Pagerank

To further elucidate the Modeling Framework, we will analyze the behavior given by choosing Pagerank as generalized fitness, detailed in Appendix B.7,

$$\mathcal{P}(i) = pr_i = \alpha \sum_{j \in \Gamma(i)} \frac{pr_j}{k_j^{out}} + \frac{1-\alpha}{N}, \tag{5.47}$$

where $\Gamma(i)$ is the set of labels of the neighboring nodes of i. Again, a bow-tie emerges in the network formation for $P_n < 1$. Now the functional P_d-dependence of the components size is more complex, as seen in Fig. 5.6. The size of the SCC is now also variable and the asymmetry of the IN and OUT sizes can be explained as follows. For $P_d = 0$ every new node is pointed to by the existing nodes. This results in $k^{in} \equiv m$. By inspecting Eq. (5.47) it is clear that the Pagerank value of a node i depends on the number of neighboring nodes pointing to it. This means that it is very sensitive to the distribution of k^{in}, which, if it has a constant value for

all nodes, results in a close-to-uniform preferential-attachment mechanism because all nodes have similar Pagerank values. As a consequence, the network is not scale-free anymore and the distribution of k^{out} can be approximated by an exponentially decaying function. In summary, this model's domain of validity is restricted to values of P_d larger than zero and the impact of varying the parameter P_d does not result in a symmetrical behavior as in the degree-dependent case seen in Fig. 5.5. This also means that this incarnation of the Modeling Framework cannot yield small IN sizes. In particular, small SCCs are associated with small OUT and large IN sizes. Similarly to the degree-dependent case, setting P_n to lower values increases the SCC size at the cost of the other components size.

$\mathcal{F}$ as Network Control or Integrated Control

As a last fitness, we test our Modeling Framework employing the centrality measures inspired by measuring the flow of control in ownership networks (see Chap. 2.9). We will use the network value as preferential attachment probability

$$\mathcal{P} = v^{\text{net}} = (I - W)^{-1} v, \tag{5.48}$$

as seen in Eq. (2.46). v^{net} can also be interpreted as network control using a linear model to estimate control (see Sect. 2.7. In addition, we also employ the integrated value as probability

$$\mathcal{P} = \tilde{v}^{\text{int}} = (I - W)^{-1} W v, \tag{5.49}$$

given in Eq. (2.42). Again, this measure can also be interpreted as integrated control using the linear model for direct control. Note that if the computation of $(I - W)^{-1}$ happens to be singular in the network evolution (see a discussed in Sect. 2.2.2) we multiply W with a dampening term similar to the one used in Pagerank in Eq. (5.47), with a value of $\alpha = 0.95$: $v^{\text{net}} = \alpha W v^{\text{net}} + v$, and $\tilde{v}^{\text{int}} = \alpha W(\tilde{v}^{\text{int}} + v)$.

As we are interested in the basic behavior of the resulting network evolution, we do not consider the corresponding corrected values $\bar{v}^{\text{net}}$, defined in Eq. (2.77) and $\bar{v}^{\text{int}}$, seen in Eq. (2.81). This is kept for future work. Especially as the computational requirements for these corrected centralities are very demanding. However, a preliminary analysis of their behavior revealed that the results are comparable to the uncorrected case. It is an interesting fact that the network evolution is not extremely sensitive to the corrections given in Sect. 2.3.4, whereas the computation of the centralities is.

Figure 5.7 displays the link-direction dependence of the model. The left diagram shows the bow-tie components size for varying P_d for v^{net} as fitness. The right plot displays this for $\tilde{v}^{\text{int}}$. The first thing that becomes apparent is that the SCC's dependence on P_d is inverted to the case employing Pagerank, as seen in Fig. 5.6. In contrast, the qualitative behavior of the IN and OUT sizes stays comparable. This observation already allows the conclusion, that this embodiment of the Modeling Framework is suited to yield networks with small IN and small SCC sizes. In effect,

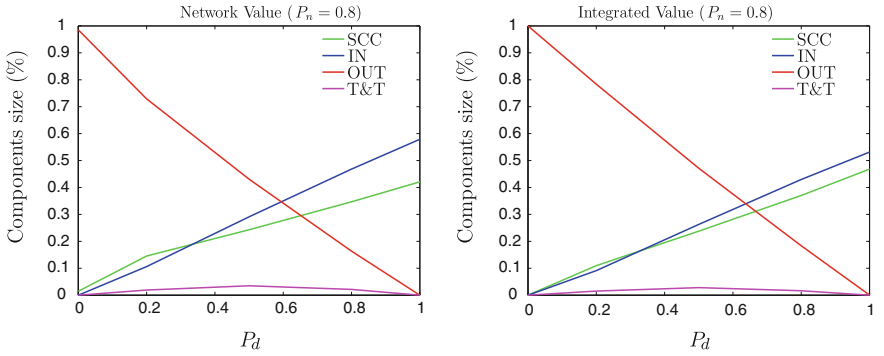

Fig. 5.7 Bow-tie components size as functions of the link direction P_d for the Modeling Framework with $P_n = 0.8$ and $m = 2$; (*left*) for the network value centrality v^{net} as fitness; (*right*) for the integrated value centrality $\tilde{v}^{\text{int}}$ as fitness; the values are averages over ten network realizations each containing 10000 nodes

network control or integrated control, interpreted as novel centrality measures, are sensitive to k^{out} due to the fact that the flow of control is against the direction of the links. Very encouraging is that the resulting parameter space of this model encompasses the distinct bow-tie signature observed in ownership networks. Again, as in the two alternative cases discussed above, lowering the value of P_n increases the overall size of the SCC at the cost of the IN and OUT.

It should be pointed out that the Modeling Framework using $\tilde{v}^{\text{int}}$ as fitness has some peculiarities. These are due to the fact that nodes i can have $\tilde{v}_i^{\text{int}} = 0$ resulting in a zero probability of ever being chosen in the preferential-attachment mechanism during the whole formation process of the network. As an example, for $(P_n, P_d) = (1, 0)$, where all new nodes get pointed to by the existing ones, every added node has $\tilde{v}_i^{\text{int}} = 0$ because there is no inflow of control ($k_i^{out} = 0$). This is a very pathological case. On the other hand, for $(P_n, P_d) = (1, 1)$ all new nodes have $k^{out} \equiv m$, resulting in a close-to-uniform distribution of control. Similarly to the Pagerank case for $(P_n, P_d) = (1, 0)$, the resulting network is not scale-free and k^{out} can be approximated by an exponentially decaying function. This reduces the validity of the model to a restricted range of P_d between zero and one, centered around the value of one half.

A final observation is that because $v^{\text{net}} = \tilde{v}^{\text{int}} + v$, this can be interpreted as in the case of the degree-dependent model seen in Eq. (5.46): the preferential-attachment probability $\mathcal{P}(i)$ is augmented by a value v_i.

Overview

To summarize, the parameter P_n is instrumental for the emergence of the bow-tie structure, as it is responsible for the degree the existing high-fitness nodes become connected among each other. Tuning its value down from unity increases the resulting

SCC's size. By tweaking P_d, one can determine the size of the IN and OUT sections. For instance, a value close to one produces a huge IN and a small OUT. This is because each new node in the network evolution then points to the existing nodes, which are high-fitness nodes, i.e., either SCC-nodes (their number depending on P_n) or IN-nodes, which increases the IN size. Only very few OUT-nodes exist, which formed when the SCC emerged early in the formation process. The T&T appear with a low probability when the new nodes attach to OUT-nodes.

5.3 The Bow-Tie Model of Ownership Networks

Now that we have come to terms with the intricacies of the Modeling Framework, the final task is to reproduce some of the empirical properties of the TNC network in the network evolution. This requires two preliminary selections:

1. What empirical features of the TNC network should be captured?
2. Which fitness-dependence should be plugged into the Modeling Framework?

In a final step, embedding the ownership network in the real-world context it pertains to, requires the modeling parameters and mechanisms to be interpreted from an economics point-of-view. This represents the first attempt in the empirical validation of the bow-tie model of ownership networks.

5.3.1 Modeling Preliminaries

Empirical Properties

Of the multitude of empirical properties discussed in Sect. 4.2, we focus on the two succinct characteristics of the TNC network:

1. its scale-free nature;
2. the existence of a highly concentrated bow-tie structure with the tiny SCC being comprised of nodes with high network control.

The empirical signature uncovered in Chap. 4 is summarized in Table 5.2. Recall from Sect. 4.2.2 that the data on k^{in} is assumed to be impacted by a systematic bias which is reflected in its peculiar distribution, seen in Fig. 4.3a. For this reason we do not attempt to reproduce the distribution of k^{in}.
It is also worth mentioning that the concept of *stylized facts* is not really applicable in our case. As we only have a single-snapshot realization of the TNC network we cannot be too confident in claiming that the measured numerical values are robust enough to represent universal empirical features, i.e., stylized facts. We are, however, confident that qualitatively the tiny SCC comprised of important economic

Table 5.2 Empirical signature of the LCC in the TNC; for the bow-tie components the values represent percentages with respect to the LCC size (463006 nodes); α_{out} refers to the estimated scaling-law exponent of k^{out}; in the last column the average total degree is shown

IN	SCC	OUT	T&T	α_{out}	$\langle k^{tot} \rangle$
1.213	0.285	70.099	28.403	−2.15	3.358

actors, the small IN and huge OUT reflect a very important feature of the TNC network.

Choice of Fitness

Concerning the choice of generalized fitness, we will focus on network control (corresponding to network value v^{net} using a linear model of control). We do not consider $\tilde{\nu}^{\text{int}}$ as it has undesirable pathologies, as discussed in Sect. 5.2.3. The GPAMs are also not considered, as we do not want a degree-driven network evolution (see more details in Sect. 5.3.4 below). Finally, as also discussed in Sect. 5.2.3, Pagerank can be understood as a diametrically opposing centrality measure to network control.

5.3.2 Scanning the Parameter-Space

From the left-hand "phase diagram" seen in Fig. 5.7 it is apparent that the range of interesting parameters are given for P_n close to one and P_d close to zero. Table 5.3 shows a scan of the relevant parameter-space. Note that an SCC size of 0.285 % corresponds to only 28 nodes in a 10000-node network. This represents quite a challenge in a network evolution model. As we are not confident how universal this value is, we are content to reproduce small SCC sizes around 1 % in the final model.

5.3.3 The Bow-Tie Model of the TNC Network

We are able to qualitatively mimic the TNC network's empirical signature with the Modeling Framework employing network control as fitness for the parameter values $P_n = 0.99$, $P_d = 0.1$ and $m = 2$. In Table 5.4 the results are summarized and compared with the empirical properties. Figure 5.8 shows the resulting degree distribution of k^{out} which can be fitted by a scaling law with an exponent $\alpha \approx -2$. In Fig. 5.9 four snapshots of the network formation are shown. The initial seeding network is comprised of nodes 1 and 2 with the link $1 \rightarrow 2$. The network at time

Table 5.3 Averaged statistics over ten network realizations, each resulting network having $N = 10000$ nodes

P_n	P_d	$\langle k^{tot} \rangle$	α_{out}	# Links	% LSCC	% IN	% OUT	% T&T
0.980	0.06	4.040	1.924	20199.4	2.318	4.123	87.313	6.246
	0.07	4.040	1.922	20199.4	2.247	4.970	85.292	7.491
	0.08	4.040	1.970	20199.4	2.647	5.662	84.252	7.439
	0.09	4.040	2.000	20199.4	2.561	6.408	82.320	8.711
	0.10	4.040	2.265	20199.4	2.766	7.333	79.988	9.913
	0.11	4.040	2.438	20199.4	2.798	8.144	78.863	10.195
0.985	0.06	4.030	1.930	20150.4	2.005	4.244	86.647	7.104
	0.07	4.030	1.930	20150.4	1.945	5.104	84.230	8.721
	0.08	4.030	1.926	20150.4	2.228	5.802	82.743	9.227
	0.09	4.030	2.083	20150.4	2.199	6.602	80.652	10.547
	0.10	4.030	2.119	20150.4	2.281	7.430	79.302	10.987
	0.11	4.030	2.381	20150.4	2.209	8.424	76.637	12.730
0.990	0.06	4.020	1.938	20099.4	1.503	4.433	83.818	10.246
	0.07	4.020	2.005	20099.4	1.540	5.243	81.999	11.218
	0.08	4.020	1.928	20099.4	1.579	6.069	79.950	12.402
	0.09	4.020	1.928	20099.4	1.694	6.873	78.562	12.871
	0.10	4.020	2.283	20099.4	1.635	7.843	75.205	15.317
	0.11	4.020	2.234	20099.4	1.738	8.721	73.655	15.886

Table 5.4 Empirical signature of the LCC seen in Table 5.2 and the best-fit achieved with the bow-tie model using the parameters $N = 25000$, $m = 2$, $P_n = 0.99$ and $P_d = 0.1$; the resulting network has 50265 links

	IN	SCC	OUT	T&T	α_{out}	$\langle k^{tot} \rangle$
TNC	1.213	0.285	70.099	28.403	−2.15	3.358
Bow-tie model	7.720	1.344	73.840	17.096	−1.922	4.021

$t = 4$ is shown in the top panel, where the new node and its links are seen in red. In the left-hand network given in the middle panel we see the emergence of the bow tie. For the first time in this network evolution a link is added between two existing nodes resulting in the formation of an SCC of three nodes, seen on the right-hand diagram in black. Finally, on the bottom panel the network comprised of 40 nodes is shown. It is clear, that for this network the high-control nodes (as indicated by the node sizes) are located in the SCC and the IN section. This is in full correspondence with the TNC network.

Recalling that $v^{\text{net}} = \tilde{v}^{\text{int}} + v$, it is an interesting observation that the choice of distribution of v has a small impact on the model's signature. For instance, choosing a log-normal distribution or a uniform one yields similar results. As a final remark, the choice of the initial seeding network does not change the qualitative outcome of the model.

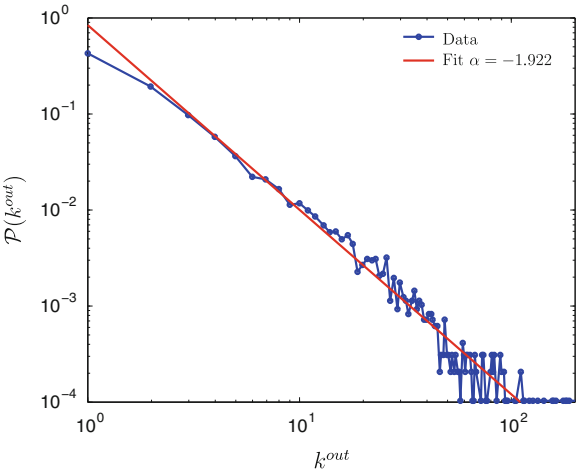

Fig. 5.8 Degree probability distribution of k^{out} for the bow-tie model, i.e., using the Modeling Framework with v^{net} as centrality, $N = 25000$ and $m = 2$; a scaling law with exponent $\alpha = -1.922$ is shown, fitted with a cutoff at $k^{out} > 55$, $R^2 = 0.9638$

5.3.4 Summary and Conclusion: The Economics Interpretation

The investigation of large-scale economic networks has only slowly started to gain the interest of academics (see also Sect. 1.1.2). The particular challenges of this line of research were recently summarized and discussed (Schweitzer et al. 2009, 2010). In this thesis we uncovered an unsuspected regularity in the organizational structure of the global TNC network: the emergence of a tiny core of interconnected key economic actors, forming an SCC and hence yielding a bow-tie topology. This is in contrast to the few other real-world complex networks exhibiting bow-tie structures where the SCCs are very large (e.g., the WWW and Wikipedia).

From a theoretical modeling perspective, what does this macroscopic feature imply for the microscopic behavior of the actors, reflected in this network structure? In general, most existing models of network evolution are perceived as being too "mechanistic" to be able to capture the true dynamical behavior of nodes and links. Moreover, the development of new dynamic models is understood as a major challenge (Schweitzer et al. 2010).

In static models of network games, a relationship between an economic agent's utility and its network centrality has been shown to exist (Ballester et al. 2006). In detail, the centrality measure of an agent gives it an objective utility, which the agent tries to maximize. In other words, there exists economic situations in which the agents try and maximize their centrality in the network.

Recall that the novel measure of network control is tailored to reflect the importance of economic actors in terms of potential power, see Sect. 6.1.3. It has a precise interpretation as control gained via ownership relations in a directed and weighted ownership network with economic values assigned to the nodes. Crucially, network

control is akin to variants of eigenvector centrality, see Sect. 2.6. It is therefore an ideal candidate to act as the utility function that the economic actors are trying to optimize. We therefore conjecture that in ownership networks the agents are striving to maximize their level of network control.

The Model

The network Modeling Framework introduced above is then an ideal toolkit to consider such microscopic dynamics, as it is conceived as a generic node/link formation model with a preferential-attachment mechanism based on some generalized fitness. This also means that the network evolution model we propose considers all three levels of analysis (see Sect. 1.1.1): node heterogeneity next to directed and weighted links.

By plugging in the network control as fitness, we are able to find a close correspondence between the simulated network signature and the TNC network's empirical one, as seen in Table 5.4. The agreement of the modelled structure with the empirical structure occurs in the following range of the parameters characterizing the link-node formation process: P_n close to one, P_d around 0.1 and $m = 2$. Recall that $P_n = 1$ is a pure node-growth model with no new links being formed between existing nodes and that $P_d = 0$ means that all the new nodes being added to the network get pointed to by the existing ones.

In summary, the network evolution is mostly driven by growth. Most new firms entering the market get owned by the most powerful existing companies. Rarely a new firm gets to own shares in one of the existing high-control corporations, however, instantaneously giving it also a high level of control. As a result, the oldest firms tend to be the most important ones. In addition, with a small but not negligible probability, new links are formed among the most powerful existing firms, resulting in the emergence of the tightly-knit core of influential firms.

This leads to some interesting and falsifiable predictions. The premise of such a high level of growth is contestable. To what extent it is feasible that new firms enter the market at a high rate and how this is applicable for TNCs should be investigated in future studies. At first blush it appears to be a modeling artefact necessary to keep the SCC small.

However, the idea that the oldest companies are the most important ones is very intuitive, as they have had more time to establish themselves in the market. That most new firms have a propensity to be owned by the existing powerful corporations is also not very controversial. If economic agents are indeed optimizing their network control, an easy way to do this, next to establishing links amongst themselves, is to prey for the new firms entering the market.

As seen in the bottom panel of Fig. 5.9, the resulting distribution of high-centrality nodes in the simulated network corresponds to the qualitative distribution in the TNC

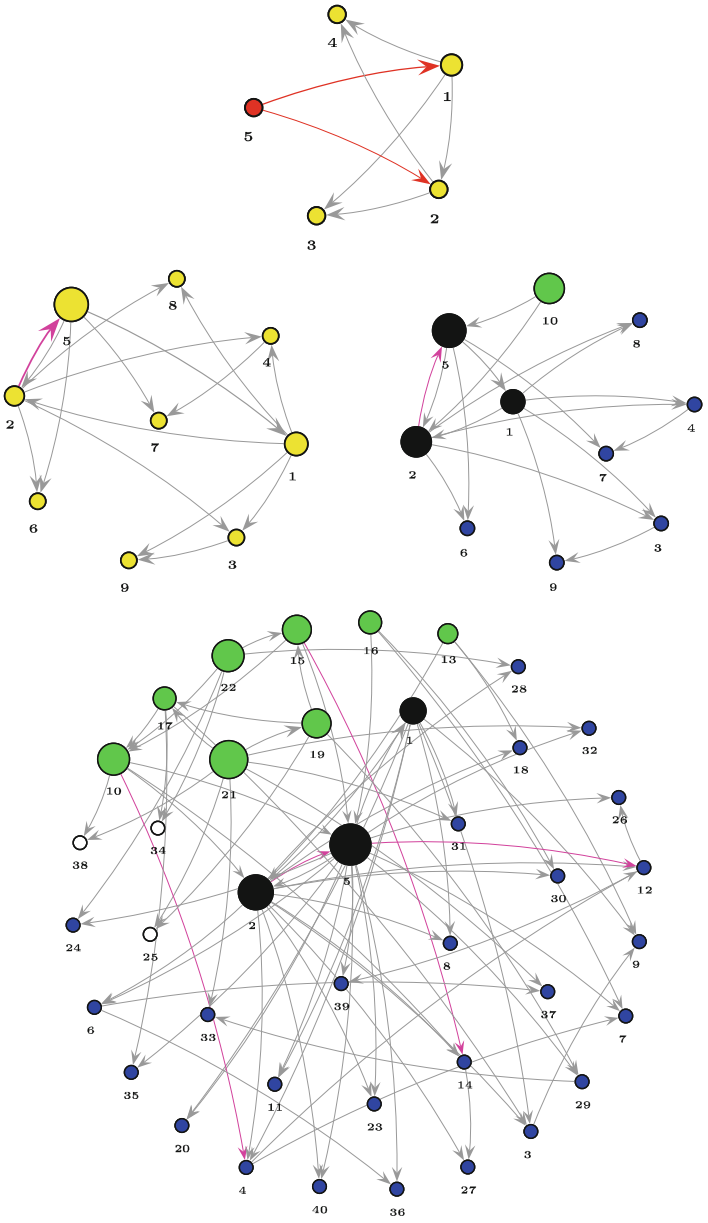

Fig. 5.9 Various stages in the network evolution; *yellow* nodes denote the network state before the bow-tie emerges; new links between existing nodes are colored *pink*; new nodes are colored *red* with m new *red* links; the bow-tie nodes are represented by *green*, *black*, *blue* and *white* nodes, corresponding to the IN, SCC, OUT and T&T, respectively; the size of the nodes reflects their network value v^{net}; the link thickness represents the weight; the parameter values are $P_n = 0.89$, $P_d = 0.235$ and $m = 2$; see the discussion in the main text (the snapshots were produced using the Cuttlefish Workbench, an open-source project found at http://cuttlefish.sourceforge.net/)

network. Next to the tiny SCC being comprised of the most powerful economic actors, the small IN has, by construction (as network control flows against the direction of links), also high-degree nodes.

In choosing $m = 2$, the number of links associated with every new node, we are able to get the average total degree to correspond to the low value of the TNC network, as seen in Table 5.4. To what degree this strong constraint reflects a true market feature is not clear. In future work it would be advisable to let the number m be probabilistic at every step in the network evolution as well, as suggested in (Giammatteo et al. 2010). Such a level of flexibility is probably more in-line with the actual mechanisms driving the evolution of the TNC network.

Another strong constraint imposed by our model is the fact that the weights of the links have no diversity. At each step in the network-formation process, the distribution of weights is simply determined by normalization, i.e., $\sum_i W_{ij} = 1$, recalling Eq. (1.1). Hence, every node i has incoming links of equal weight. This is in contrast to the empirical distribution of the link weights. This constraint could be eased in future versions of the model. However, the complexity of having varying weights included in the dynamics is a non-trivial task (Barrat et al. 2004a,b).

Finally, it should be investigated why the corrected control centralities do not yield networks with a better correspondence to the TNC network and why the distribution of v appears to play a marginal role in the network evolution.

In Closing

These results shed new light on the formation of networks of economic agents. They show that such networks are highly non-random and that economic forces are heavily shaping their evolution. In particular, the observed structures are explained by a strong reinforcement mechanism that gives priority of action to the powerful actors.

Because our simple model has very few underlying assumptions (only link and network growth next to a centrality-driven preferential attachment mechanism), this endeavor can be seen as the beginning of an extensive modeling effort of ownership networks. There are many conceivable extensions, refinements and additions possible. However, as we are able to reproduce the key bow-tie signature next to the empirical out-degree scaling-law exponent of the TNC network only with such a simple model and network control as centrality, which is in-line with the economic intuition (Ballester et al. 2006), we are confident to have uncovered a genuine micro-foundation for the empirical pattern that is not an arbitrary modeling artefact.

In essence, the model we present here represents a minimal setup able to capture the TNC network's basic characteristics. It is highly motivating for future work, that this is even possible with such a crude model. Although the various refinements proposed

here probably would result in a better correspondence and fewer modeling artefacts, the main result we show here are the following:

There exists a generic framework for simulating the network evolution which can reproduce the various bow-tie structures observed in real-world complex networks.

Applied to the TNC network, it is possible to empirically validate the model to an encouraging degree.

References

C. Ballester, A. Calvo-Armengol, Y. Zenou, Who's who in networks. Wanted: the key player. Econometrica, **74**(5):1403–1417 (2006)

A.L. Barabási, R. Albert, Emergence of scaling in random networks. Science **286**, 509 (1999)

A. Barrat, M. Barthélemy, A. Vespignani, Weighted evolving networks: coupling topology and weight dynamics. Phys. Rev. Lett. **92**(22), 228701 (2004a)

A. Barrat, M. Barthélemy, A. Vespignani, Modeling the evolution of weighted networks. Phys. Rev. E **70**(6), 66149 (2004b)

M. Boguñá, R. Pastor-Satorras, Class of correlated random networks with hidden variables. Phys. Rev. E **68**(3), 36112 (2003)

A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, S. Stata, A. Tomkins, J. Wiener, Graph structure in the web. Comput. Netw. **33**, 309 (2000)

G. Caldarelli, A. Capocci, P. de Los Rios, M.A. Muñoz, Scale-free networks from varying vertex intrinsic fitness. Phys. Rev. Lett. **89**(22), 258702 (2002)

A. Capocci, V.D.P. Servedio, F. Colaiori, L.S. Buriol, D. Donato, S. Leonardi, G. Caldarelli, Preferential attachment in the growth of social networks: the internet encyclopedia wikipedia. Phys. Rev. E **74**(3), 36116 (2006)

S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, E. Shir, A model of internet topology using k-shell decomposition. Proc. Natl. Acad. Sci. **104**(27), 11150 (2007)

D. Donato, S. Leonardi, S. Millozzi, P. Tsaparas, Mining the inner structure of the web graph. J. Phy. A Math. Theor. **41**(22), 224017 (2008)

S.N. Dorogovtsev, J.F.F. Mendes. Evolution of networks with aging of sites. Phys. Rev. E **62**(2), 1842–1845 (2000a)

S.N. Dorogovtsev, J.F.F. Mendes, Scaling behaviour of developing and decaying networks. EPL **52**, 33–39 (2000b)

S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003)

S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Structure of growing networks with preferential linking. Phy. Rev. Lett. **85**(21), 4633–4636 (2000)

S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Giant strongly connected component of directed networks. Phys. Rev. E **64**(2), 25101 (2001)

R.A. Durrett, *Probability Theory and Examples* (Thomson Learning, London, 2004)

D. Garlaschelli, A. Capocci, G. Caldarelli, Self-organized network evolution coupled to extremal dynamics. Nat. Phys. **3**(11), 813–817 (2007)

P. Giammatteo, D. Donato, V. Zlatić, G. Caldarelli, Pagerank Based Preferential Attachment Model for the Evolution of the WWW (Forthcoming; privately obtained from G, Caldarelli, 2010)

J.B. Glattfelder, S. Battiston, Backbone of complex networks of corporations: the flow of control. Phys. Rev. E **80**(3), 36104 (2009)

P.L. Krapivsky, S. Redner, Organization of growing random networks. Phys. Rev. E **63**(6), 66123 (2001)

P.L. Krapivsky, G.J. Rodgers, S. Redner, Degree distributions of growing networks. Phys. Rev. Lett. **86**(23), 5401–5404 (2001)
M.E.J. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E **64**(2), 26118 (2001)
S.M. Ross, *Introduction to Probability Models* (Academic Press, San Diego, 2006)
F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D.R. White, Economic networks: the new challenges. Science **325**(5939), 422 (2009)
F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, D.R. White, Economic Networks: What Do We Know and What Do We Need to Know? (Forthcoming in, Advances in Complex Systems, 2010)
V.D.P. Servedio, P. Buttà, and G. Caldarelli. Vertex intrinsic fitness: how to produce arbitrary scale-free networks. Phys. Rev. E, **70**, 056126 (2004)
G. Siganos, S.L. Tauro, M. Faloutsos, Jellyfish: a conceptual model for the AS onternet topology. J. Commun. Netw. **8**(3), 339 (2006)
B. Söderberg, General formalism for inhomogeneous random graphs. Phys. Rev. E **66**(6), 66121 (2002)
B. Tadić, Dynamics of dDrected graphs: the world-wide web. Physica A **293**(1–2), 273–284 (2001)
A. Vázquez, M. Boguná, Y. Moreno, R. Pastor-Satorras, A. Vespignani, Topology and correlations in structured scale-free networks. Phys. Rev. E **67**(4), 46111 (2003)
S. Vitali, J.B. Glattfelder, S. Battiston, The network of global corporate control. PloS ONE **6**(10), e25995 (2011)
S. Wolfram, *A New Kind of Science* (Wolfram Media, Champaign, 2002)
S. Zhou, R.J. Mondragón, Accurately modeling the internet topology. Phys. Rev. E **70**(6), 66108 (2004)