Physica A 389 (2010) 3661–3673

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Analysis and models of bilateral investment treaties using a social networks approach

Daniela Saban [a], Flavia Bonomo [a,b], Nicolás E. Stier-Moses [c,*]

[a] Computer Science Department, FCEN, University of Buenos Aires, Buenos Aires, Argentina
[b] CONICET, Argentina
[c] Graduate School of Business, Columbia University, New York, NY 10027, USA

ARTICLE INFO

Article history:
Received 8 January 2010
Available online 11 April 2010

Keywords:
Bilateral investment treaties
Complex networks
Network growth models
Social networks

ABSTRACT

Bilateral investment treaties (BITs) are agreements between two countries for the reciprocal encouragement, promotion and protection of investments in each other's territories by companies based in either country. Germany and Pakistan signed the first BIT in 1959 and since then, BITs are one of the most popular and widespread form of international agreement. In this work we study the proliferation of BITs using a social networks approach. We propose a network growth model that dynamically replicates the empirical topological characteristics of the BIT network.

1. Introduction

Bilateral investment treaties (BITs) are agreements between two countries for the reciprocal encouragement, promotion and protection of investments in each other's territories by companies based in either country. The signing of the first BIT between Germany and Pakistan in 1959 initiated the creation of a network of treaties that has experienced continuous growth. By the end of 2005, 179 countries (out of approximately 200) had signed at least one BIT and there were a total of 2460 BITs in force. Nevertheless, although a significant number of new BITs were signed in the first half of the 1990s, the rate at which new BITs were signed started to decrease afterwards. Indeed, the number of new treaties in 2005 was down 60% compared to 1995. Despite this decrease, BITs are still one of the most popular and widespread forms of international treaty.

There have been many studies aimed at understanding the proliferation dynamics of BITs. Their results are inconclusive and there is a large debate about the reasons why countries sign BITs and the effects caused by their signing [1–4]. Many of these studies assume that the motivation to sign BITs is to improve the chances of receiving foreign direct investment (FDI), although there is mixed evidence in the literature regarding this assumption. In particular, Neumayer and Spess discuss that a higher number of BITs raise the FDI that flows to a developing country [3], and Elkins, Guzman and Simmons argue that the spread of BITs is driven by international competition among potential host countries for FDI [2].

This multi-disciplinary work departs from earlier approaches and studies the body of BITs using a *complex social networks* perspective. A social network is a structure made of nodes that are tied by one or more specific types of interdependency. In our case, nodes represent countries and a tie between nodes indicates the existence of a BIT. Networked systems from the real world have routinely been studied using this perspective. Examples include the Internet [5], the World Wide Web [6], scientific collaboration networks [7], and metabolic networks [8]. These networks are referred to as *complex* because they have a large number of nodes that are connected forming non-trivial topological features. The mentioned connection

* Corresponding author.
E-mail addresses: dsaban@dc.uba.ar (D. Saban), fbonomo@dc.uba.ar (F. Bonomo), stier@gsb.columbia.edu (N.E. Stier-Moses).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2010.04.001

patterns are neither purely regular nor purely random. Since the 1950s, several random network growth processes have been proposed with the goal of emulating the topology of a complex network. The first of these is the random graph model of Erdős and Rényi, which connects every pair of nodes uniformly and independently with probability p [9]. Running this process on n nodes creates a graph $ER(n, p)$ that has approximately $pn(n - 1)/2$ edges distributed randomly. With the advent of more powerful computers, the empirical analysis of large, real-world networks has shown that many of them share some fundamental structural properties, such as small-world effect, high clustering coefficient and power-law degree distribution. With these results in hand, scientists started to question whether *ER* graphs gave rise to networks similar to those generated by real-world complex systems.

The study of the social network generated by the BITs departs from related studies in the literature for the following main reasons. First, the full history of the network is compact and available to us. While most works on real-world networks study the network evolution process over a short period of time, typically no more than 10 years, we will study the evolution process from birth to actuality, covering a period of 45 years. Second, while most networks studied in the complex networks literature are large and sparse with potentially infinite growth [10,6,5], the BIT network is small and dense, and has almost reached its limit of growth at node level. The fact that the network is small will allow us to study the local properties of the network such as its cohesiveness (as given by the clique or quasi-clique numbers). An analysis of these properties is not present in most of the existing literature. Third, traditional papers from the literature of social networks view them as static graphs, and concentrate their attention on the analysis of structural properties of snapshots at different times. We will study the BITs using a dynamic perspective, paying special attention to growth processes that generate networks with similar properties [7]. The main two conclusions that can be drawn from our study are that a network growth process based on a combination of preferential attachment and the fitness model is a good fit for the BIT network, and that the reason why less countries signed new BITs in the period 1995–2005 is the existence of some saturation whereby countries had already signed the BITs that were most important to them.

There are other networks representing an interaction between countries that have been studied from a social networks perspective. To cite the most relevant, [11] studies the topology of the *world trade web*, defined by international import/export trade relationships. In follow up work, [12] focuses on a directed version of the network and looks at its evolution. The paper [13] explores the complex relationships between countries in the *Eurovision Song Contest*, by creating a dynamic network from voting data over a ten-year period. The evolution in both of these application domains allows for the relationships to change arbitrarily over time. This means that edges could be added or deleted from one year to the next. On the contrary, our network only admits the addition of edges: once a BIT is signed, it remains signed forever.

This paper is organized as follows. Section 2 starts by describing the dataset and laying out the groundwork by reporting on the structural properties of the BIT network. Section 3 discusses the evolution of BITs over time and explains the difference between the BIT network and the most-commonly studied big and sparse networks. In Section 4, we propose the models that capture the main aspects of the BIT network and measure the goodness-of-fit using topological characteristics. Finally, we conclude in Section 5 with some opportunities for further work.

2. The BIT network

We used a dataset collected by the United Nations Conference on Trade and Development (UNCTAD) [14,15]. It contains all BITs that were signed starting with the first BIT in 1959, up to 2005.[1] The set of pairs of countries that signed BITs can be regarded as a social network, where the countries are the nodes and an edge between two countries is present if they signed a treaty. In some limited number of cases, a dyad of countries signed a BIT more than once. We only consider the oldest treaty when that happens because the new one is usually a revision and a ratification of the BIT. In addition, some countries like Czechoslovakia and Yugoslavia have divided, so they stopped to exist as countries. Since this is a second-order consideration because it is not a frequent event, we consider that the network growth process is monotone and, thus, we never delete a country or an edge from our network. Consequently, we treat newly formed countries such as the Czech Republic and Slovakia as new countries that join the network. Overall, our network contains 2460 treaties signed by 179 different countries.

To study the dynamics of the BIT network, we define N_y for each year $y \in \{1959, \dots, 2005\}$ as the set of countries that signed at least one treaty before or in year y, and E_y as the set of treaties signed before or in that year. We let $\mathrm{BIT}_y = (N_y, E_y)$ be the state of the network at year y. Fig. 1 summarizes its growth by plotting the number of new countries and new treaties per year, in the period 1960–2005.

2.1. Structural properties

We now study the properties of the BIT network, focusing on the evolution of these properties over time. Comparing our measurements with what is expected for an *ER* graph of similar size, we start to uncover significant patterns in the network. To obtain representative measurements for random graphs, we average results over 50 independent trials. This section often uses graph terminology; readers are referred to [16] for an introduction to graph theory.

[1] For each BIT, the dataset includes the signature date and the date of entry into force. We just consider the year when the BIT was signed.

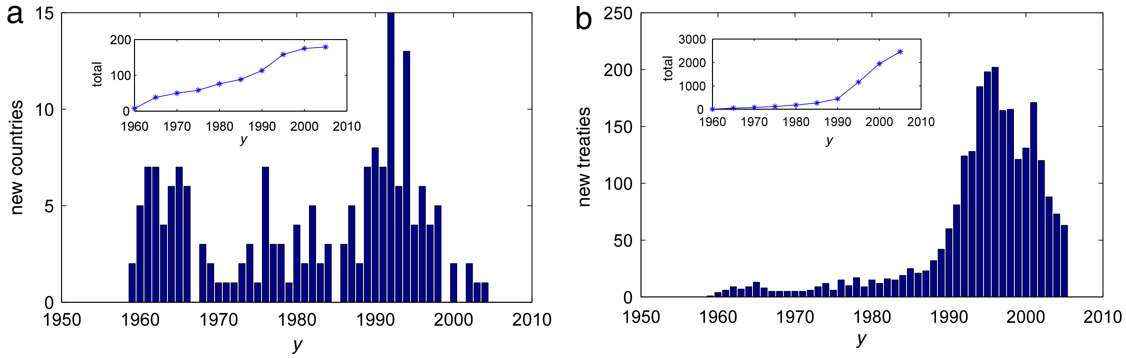

Fig. 1. (a) New countries per year. (b) New treaties per year. *Note*: The insets show cumulative numbers.

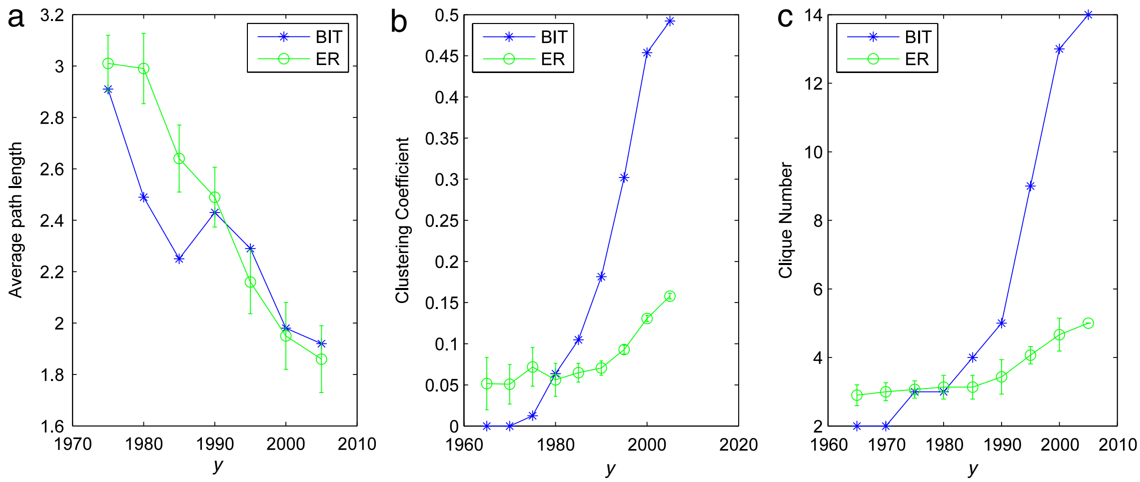

Fig. 2. Comparison of the BIT network and *ER* graphs as a function of time. (a) Average path length. (b) Clustering coefficient. (c) Clique number. *Note*: The error bars correspond to 1 standard deviation above and below the curve.

Table 1
Evolution of size and density in the BIT network.

Year y	1975	1980	1985	1990	1995	2000	2005
$\lvert N_y \rvert$	58	76	88	113	158	175	179
Density	0.0695	0.0635	0.0700	0.0704	0.0936	0.1307	0.1587

The *density* of a network is defined as the number of edges $|E_y|$ it contains over the maximum number of possible edges $|N_y|(|N_y|-1)/2$. Table 1 shows this parameter in intervals of five years. The density remains constant until around 1990, but afterwards it shows an almost perfect linear growth with an increase in density of 0.006 per year ($R^2 = .993$). The observed density is significantly different from that of most real-world networks described in the literature. For example, the density of BIT_{2005} is 0.1587, while in the graphs analyzed by Refs. [10,6,7,5,8], densities were in the interval [0.00001, 0.001]. The reason of this comparatively high density is that the network is small and shows signs of saturation (see Section 3.2).

One of the most celebrated properties of social networks is the *small-world property* [17]. Although an arbitrary pair of nodes may not be connected directly, nodes are separated from each other by a small number of hops. One way to capture the interconnectedness of a network is by computing the average distance over all pairs of nodes, which we denote by ℓ. These values, plotted in Fig. 2(a), indicate that the BIT network has the small-world property. The figure also compares the values of ℓ to the average distances in a sample of *ER* random networks of the same size. (For a year y, we generated each *ER* graph in the sample using parameters $n = |N_y|$ and $p = 2|E_y|/|N_y|(|N_y| - 1)$.) With the exception of 1980 and 1985, ℓ lies within the 95% confidence intervals for the distance in random graphs (note that the figure does not display these intervals; it only shows the standard deviation), providing some support for *ER* graphs. Nonetheless, we will soon see that other parameters do not frequently lie within the corresponding confidence intervals.

Now we look at the *clustering coefficient* of a graph, which provides a local measure of density. Indeed, the clustering coefficient of a country v represents the probability that two countries that have both signed a treaty with v sign a treaty with each other. Technically, the clustering coefficient C_v of a vertex v is defined as the ratio between the total number of the edges between v's neighbors and the total number of all possible edges between them; the clustering coefficient of a graph is the average C_v among all vertices. It is easy to see that the expected value of the clustering coefficient of an $ER(n, p)$ graph is p.

Fig. 2(b) depicts the clustering coefficient for the BIT networks at different times and compares it to that of *ER* graphs. As opposed to the conclusion drawn for the small-world property, it can be observed that both networks deviate considerably. While the curve for *ER* graphs replicates the values of density described in Table 1, we can distinguish three main phases in the curve for the BIT network: the clustering coefficient of the BIT network is close to 0 until 1975 because the network is basically bipartite; the clustering coefficient increases considerably from 1976 to 1985; and the empirical curve stabilizes at roughly three times the theoretical one after 1986. This provides an indication that there was a radical change in the dyads that signed the early BITs, compared to those signed more recently. Initially BITs formed an almost bipartite network because they were signed almost exclusively between developed and developing countries, which explains the low clustering coefficient values (i.e. a bipartite graph does not contain triangles and hence the clustering coefficient is 0). The increase of the clustering coefficient in the second phase agrees with the time developing and developed countries began signing BITs within their pier groups. However, the increasing tendency of the clustering coefficients that characterize the third phase can be explained by the following two main reasons. The general increase in density alluded to in Table 1 caused the formation of many triangles, which produced an increase in the clustering coefficient of the *ER* graphs. But more specifically, in this phase there was a considerable increase in the number of BITs signed between developing countries. This created many triangles in the BIT network that consist of one developed and two developing countries, which the *ER* graph cannot reproduce because it does not consider the development level of countries.

2.2. Cohesive subgroups

A structure that proves to be specially relevant to our study is given by *cohesive subgroups*, defined as subsets of actors among whom there are relatively strong, direct, intense, frequent or positive ties [18]. In our context, they represent blocs of countries with a high density of treaties signed between them. More specifically, a possible way to characterize these blocs is by using *cliques*, which are subgraphs of the network whose nodes are fully pairwise connected. For example, the size of the largest clique – usually referred to as *clique number* – in 2005 is 14 and one arbitrarily chosen maximum clique consists of Albania, Bulgaria, China, Croatia, Egypt, Germany, Hungary, Poland, Portugal, Romania, Russian Federation, Slovenia, Turkey, and Ukraine. Fig. 2(c) plots the time-series of clique-numbers in the BIT network. We can draw a parallel to the analysis of clustering coefficients above. This provides further evidence that the formation process of the BIT network cannot be approximated accurately with *ER* graphs.

Looking at different largest cliques in the BIT network at different times, it is hard to detect a pattern among the countries forming them because countries are in different regions and have different economical levels. However, the different maximal cliques have many countries in common, possibly indicating that they are part of a larger cohesive subgroup that is not a clique because it is missing some edges. Notice that a dyad of countries that may not be willing to sign a treaty because, e.g. a long-standing conflict could still be part of the same bloc. This motivates us to consider *quasi-cliques*, defined as a subgraph with a pre-specified edge density. Formally, given a real number $0 \leq \gamma \leq 1$, a graph $G = (N, E)$ is a γ-quasi-clique if $2|E|/(|N|(|N|-1)) \geq \gamma$. Quasi-cliques can be interpreted from a bicriteria optimization point of view by considering that the size of the subgraph is the first objective and its density is the second objective. This problem is computationally hard (it is NP-hard and no constant-factor approximation algorithm can exist), but we developed an integer programming formulation and a cut generation procedure that allows us to solve the problem. We do not describe the procedure here because of lack of space; the details can be found in Ref. [19]. Using our procedure, we compute the size of the maximum γ-quasi-cliques for various values of γ and compare them to the clique number. The conclusion is that the quasi-clique number increases at a faster pace than the clique number, indicating the presence of highly cohesive subgroups with more members than those we can detect by considering regular cliques. One interesting empirical question that we leave open is to study the reasons and find the covariates that explain why some BITs were never signed.

2.3. Degree distribution

The degree of a node is defined as the number of neighbors, which corresponds to the number of BITs signed by the corresponding country. The degree distribution $\pi_y(k)$ is the probability that a node chosen uniformly at random has degree k. Note that an $ER(n, p)$ graph has a binomial degree distribution with parameters $n - 1$ and p. We study the evolution of the degree distribution of the BIT network by comparing the empirical distributions in 1975, 1990, 2000 and 2005. The interval of 15 years between those snapshots was chosen to allow the network to change and is in correspondence to the phases described previously. We also added the year 2000 because it is another recent observation and we wanted to see when the change of regime that we describe below happened.

The degree distributions corresponding to those four years are skewed to the right, in marked contrast with *ER* graphs that have lighter tails. However, the *coefficients of variation* (CV) decrease over time ($CV_{1975} = 1.55$, $CV_{1990} = 1.34$, $CV_{2000} = 1.03$, $CV_{2005} = 0.94$), implying that the tails have become lighter. As shown in Fig. 3(a) and (b), the degree distributions of the BIT network in 1975 and 1990 seem to be consistent with a power-law distribution. Instead, Fig. 3(c) and (d) show that

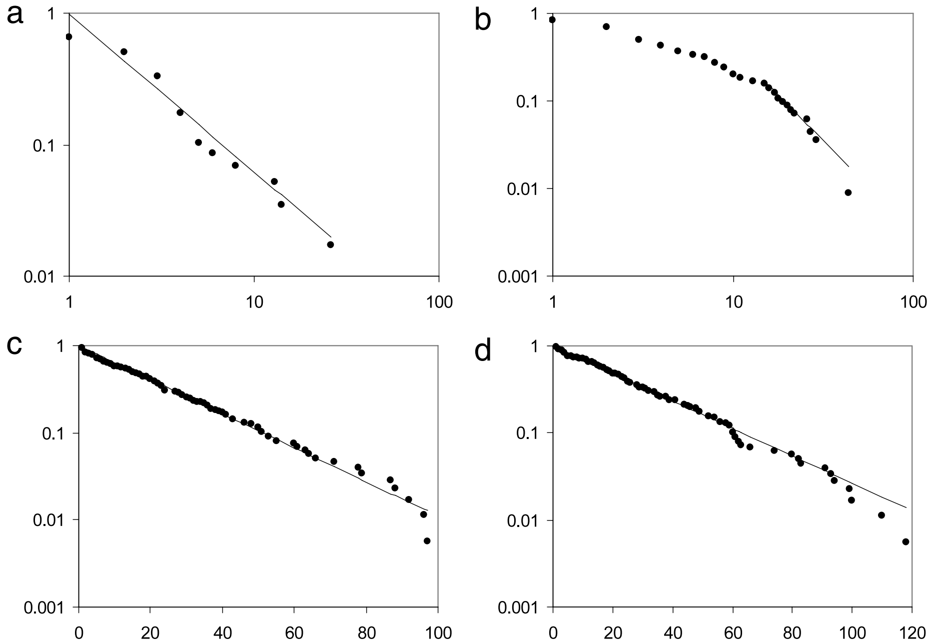

Fig. 3. Complementary cumulative distribution function of degrees for years (a) 1975, (b) 1990, (c) 2000, and (d) 2005. Note: (a) and (b) are compared to a power law distribution on a log–log scale, and (c) and (d) are compared to an exponential distribution on a log–linear scale.

the degree distributions of 2000 and 2005 fit an exponential distribution remarkably well. The transition from power law to exponential was "smooth" in the sense that the distribution in 1990 fits an exponential reasonably well although not as well as the power law. Actually, in 1990 there seem to be two different scaling regimes, exactly as discussed in Ref. [7] where it is argued that the situation can be mistaken for an exponential. In all plots but (a), the tail is thinner than should have been the case for the fitted distribution. This suggests the existence of a saturation effect whereby countries find it less beneficial to sign additional BITs because they have already signed the most important ones.

Regarding the change in the distribution that happened over the 1990s, degrees went from having the relatively fat tails of a power law to the much slimmer tails of an exponential. We believe that this change is related to the increase in density and to the finite size of the BIT network. Indeed, the network cannot maintain the degree distribution it had in its early years, including the coefficient of variation and the size of the right tail, because that would require that many countries sign too many treaties. This is unlikely, or even impossible, given that there are approximately 200 countries in the world. While the maximum number of treaties signed by a country was 67 in 1990 (with a total of 446 signed treaties at the time), it was 131 in 2005 (with a total of 2460 treaties).

3. Network growth processes

The statistical study of the properties of the BIT network performed in the previous section helps us to start understanding the dynamics of the signing of treaties. This section continues in this direction and proposes network growth processes that emulate what is observed empirically.

Real networks can grow in different ways. Namely, some grow when new nodes join the network and connect to existing nodes, which is referred as *growth at new node level*; some other networks maintain a fixed number of nodes and evolve by creating or removing edges between existing nodes, which is referred to as *growth at internal edge level*. The BIT network, as well as the Internet and the World Wide Web, present both types of growth.

A node joins the network by connecting to an already existing node. A straightforward way to model this is by assigning a *fitness* value that measures how likely an existing node is to attract new connections. Consequently, one can estimate parameters $\eta(v)$ for each node v, and link new nodes to existing ones with a probability proportional to $\eta(v)$ [20]. Unfortunately, this process does not explain why empirical degree distributions have heavy tails. *Preferential attachment* is a process that has frequently been used to explain the observed power law degree distributions. According to preferential attachment, the likelihood of a new node connecting to an existing node is proportional to its degree. Since aspects of both processes seem to be present in practice, one can combine both models, which has been referred to by *generalized preferential attachment* [20].

Although many theoretical models of social networks have previously assumed that both nodes and edges grow at a constant rate [7,8,10], Fig. 1 shows that the number of new countries and new treaties as a function of time is not constant, implying that growth is non-linear. We highlight that since the late 1990s, the number of new nodes and of new treaties started to decrease. The decrease in the number of countries that join the BIT network every year is caused by saturation since there are around 200 countries in the world and 179 of them were part of BIT network by the end of 2005. The remaining countries may not be inclined to sign BITs and hence soon there will not be more countries to add to the network. This is essentially different from social networks such as Internet and the World Wide Web, where the potential growth at node level is infinite. There is a significant debate concerning the reasons that motivated the decrease in the number of new treaties per year. Countries may have already signed all the treaties that they consider beneficial so they are running out of candidates, or countries in the last decade may have been busy dealing with other issues like wars and economic crises and therefore had less resources to devote to signing new treaties. Section 3.2 sheds some light on this issue using empirical observations, while Section 4.4 furthers the analysis using simulation.

3.1. Preferential attachment

We use the signature of BITs to determine whether preferential attachment at the new node level and at the internal edge level exist. To do so, we adapt a method proposed by Barabási et al. in Ref. [7] to work with small networks by grouping data into quartiles. We start with the effect for new nodes. If preferential attachment at the new node level is present, then the likelihood that its first BIT is signed with a given country v is proportional to the number of BITs that v has already signed. In other words, we discretize time by considering periods of one year, and sample the second country with probabilities $\Pi_y(v)$ proportional to their degrees. For a given year y, we call a node *old* if it was created earlier than y, and *new* if it joins during that year. If a node v has degree equal to $k_y(v)$ at the beginning of the year, we have that $\Pi_y(v) = k_y(v)/\sum_w k_y(w)$. New nodes connect to old ones in each period and generate $\Delta k_y(v)$ new arcs incident to each node v. According to the preferential attachment process, $k_y(v)$ and $\Delta k_y(v)$ should be strongly positively correlated.

Since our network is much smaller than the co-authorship network studied by Barabási et al., the new nodes joining the network each year may not be sufficient to observe statistically significant results when measuring relations between $k_y(v)$ and $\Delta k_y(v)$. In the history of the BIT network, between 0 and 15 countries per year joined the network, and most often it was less than 8. Since disaggregated data is not sufficient, we aggregate countries into four groups, according to the quartiles of the degree of the country it signs its first treaty with. Indeed, for a year y and an old country $v \in N_y$, we let $Q(y, v) \in \{1, \dots, 4\}$ be the quartile where $k_y(v)$ lies. For all countries v, we obtained $Q_y(v) := Q(y, v')$, where y and v' are the year when and the country with which, respectively, v signed its first BIT. We do not consider countries that joined the network before 1964 because the number of active nodes is too small for the quartiles to be significative. The histogram of $Q_y(v)$ in Fig. 4(a) shows that most new countries sign their first BIT with very active countries. This supports the idea that preferential attachment at new node level is present in the network formation process.

Now we extend the analysis by quartiles to the case of internal edges. Adapting the analysis of Barabási et al., we let $\Pi_y(q_1, q_2)$ be the probability that two countries whose degrees lie in quartiles q_1 and q_2, respectively, sign a BIT in year y. Notice that $\Pi_y(q_1, q_2) = \Pi_y(q_2, q_1)$ because the network is undirected. Fig. 4(b) depicts $\Pi_y(q_1, q_2)$ for $y = \{1975, 1990, 1996, 2005\}$. Here, we used 1996 instead of 2000 because it was the year in which the maximum number of treaties were signed. It can be observed that Π is monotone on q_1 and q_2 with the exception of $q_1 = q_2 = 4$. The monotonicity supports the existence of a preferential attachment process. The exception with the fourth quartile can be explained by the following two arguments. As we will see in Section 3.2, countries in the fourth quartile cannot or do not want to sign many more treaties because it is likely that they have already signed the BITs that are relevant to them. Hence, later in the network history it is more likely that pairs of countries that tend to sign many treaties have already signed a BIT between them. For this reason it is more likely that they sign BITs with countries in the third quartile.

This motivated us to look at an alternate definition of $\Pi_y(q_1, q_2)$ where we divided the number of BITs signed between countries with degree in quartiles q_1 and q_2 by the number of BITs that *could have been* signed between those pairs. This definition and the previous one are very similar; the difference is that the denominator before consisted of all dyads with countries in the quartiles, and the current one consists of all those that are not already signed. The modification allows us to incorporate the saturation of the network to the definition because countries that have already signed a BIT cannot sign it again. We observed that the value of $\Pi_y(4, 4)$ increased, indicating that by the time the two countries reach the fourth quartile, there is a heightened probability that they have already signed a BIT.

In a network purely subject to a preferential attachment growth process, older nodes are more likely to have higher degrees than newer ones. Indeed, an older node has a longer lifetime to accumulate edges so it is more likely to increase its degree. But it is not clear that in the BIT network older nodes have more connections. This can be appreciated in Fig. 5, which shows a boxplot for degrees in 2005 as the function of the year when countries joined the network. Besides not showing negative trend, the boxplot indicates that there is a large variability in the degrees of nodes and that countries that joined the network in the same year may end up with a very different number of treaties signed.

Regarding whether preferential attachment is present or not, we saw in Section 2.1 that the degree distribution is not always scale-free, which is in opposition with preferential attachment. Amaral et al. suggest that node aging, the cost of adding arcs and the limited capacity of a vertex may prevent the preferential attachment process from producing a scale-free network [21]. In our case, node aging is not a limiting factor since all the countries are possible candidates for signing BITs.[2]

[2] This is not entirely correct because two countries do not exist anymore so they cannot possibly sign a new BIT. Nonetheless, this represents a negligible error compared to the size of the network.

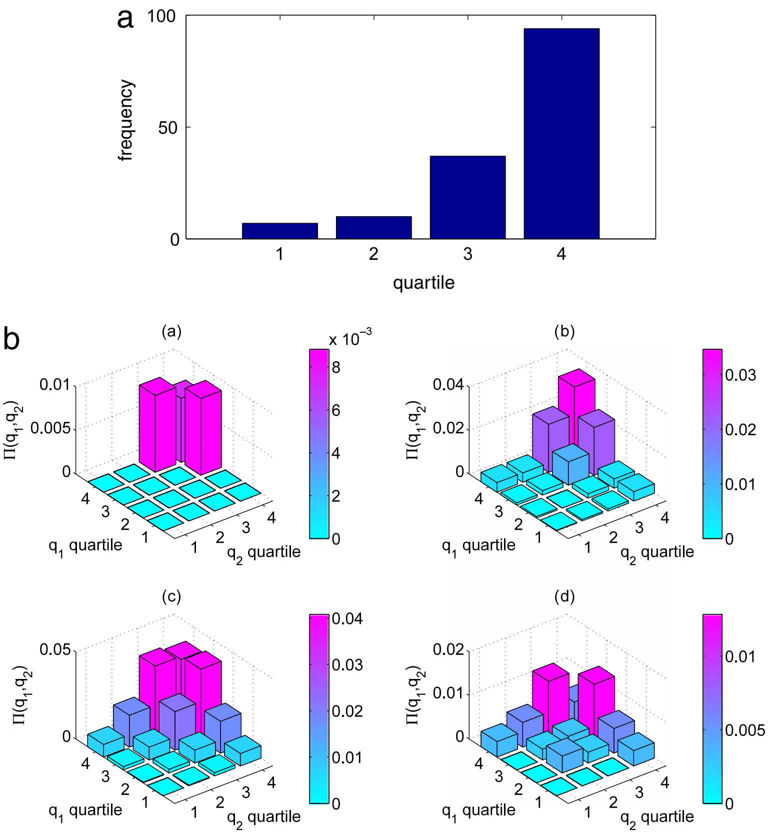

Fig. 4. (a) Histogram of $Q_y(v)$. (b) Probabilities $\Pi_y(q_1, q_2)$ for $y = \{1975, 1990, 1996, 2005\}$.

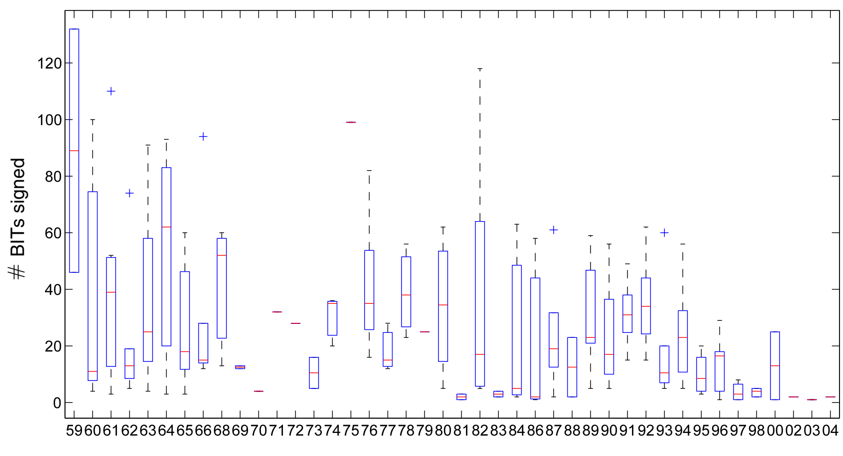

Fig. 5. Boxplot of the number of BITs signed by 2005 grouped by the year when countries joined the network.

The latter two aspects are discussed in Section 3.2. Since a preferential attachment process cannot fully explain the properties empirically observed in the BIT network, we also consider an enriched model known as generalized preferential attachment that incorporates the countries' abilities or desires to sign treaties. This involves adding a fitness parameter for every country, which allows some countries that joined the network later to "catch up" with early members. Although Fig. 4(b) does not contradict the existence of preferential attachment at the internal edge level, there is an alternative explanation to the shape of $\Pi_y(q_1, q_2)$. Nodes with a high fitness parameter are more likely to acquire more edges, regardless of their degree. Hence, the probability that two nodes with high degree connect with each other is high.

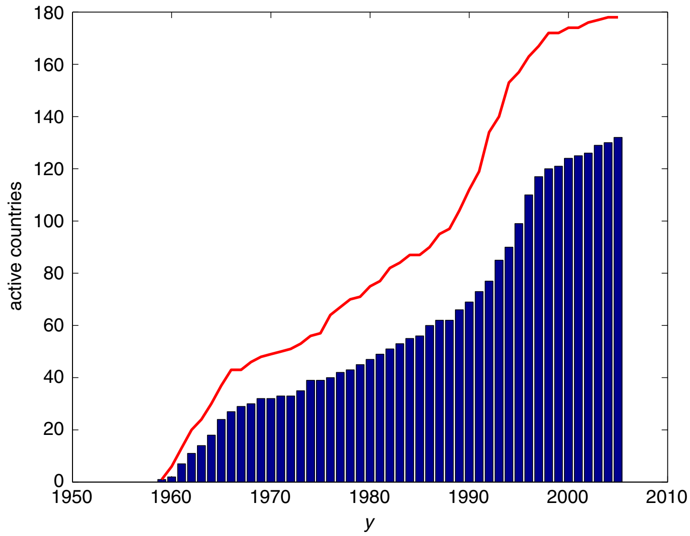

Fig. 6. Degree of Germany (bars) vs. number of active countries $|N_y|$ (line) for $y \in \{1959, \ldots, 2005\}$.

3.2. Saturation in small social networks

Nodes in the BIT network have a limited capacity because there are 200 countries in the world and hence they cannot possibly sign an unbounded number of BITs. This upper bound on the number of possible treaties has the effect of countering the effect of preferential attachment and lowering the probability that a country with high degree signs new treaties. In other examples of social networks such as the World Wide Web and the Internet, capacity is unbounded and nodes with a high degree could continue to acquire new edges indefinitely. In the scientific collaboration network, there is no upper bound to the number of papers or collaborators an author can have. Besides the saturation effects arising from the limited number of countries, another barrier that affects the signature of new BITs is given by the economic and political costs involved. BITs allow for international arbitration implying significant sovereignty costs for the countries involved, and redistribution affects the returns that the host country would have perceived. These factors can be perceived as a cost to both countries, which can be used to explain that countries only sign a BIT whenever its benefits outweigh its cost. Although a fixed cost when adding new links is present in other social networks, the significance is lower because, e.g. the cost of adding a new link to a webpage is negligible, or the cost of collaborating with a co-author that published many papers (a "star") is possibly less than that of a co-author that is inexperienced.

Out of all the countries in the evolution of the BIT network, Germany has always been the one with highest degree. This can be observed in Fig. 6, which shows the time-series of degree of Germany along with the number of active countries; there is a strong correlation between the rate at which Germany signs new treaties and the rate at which countries join the network. This suggests that Germany signs a treaty with a (fixed) fraction of the countries that join the network. Hence, the probability that Germany signs a treaty has not changed over the years. This behavior was observed for most of the countries that consistently have a high degree. Roughly, they have been observed to adhere to the following process: (1) a country joins the network, (2) it actively signs treaties until its degree is close to the number of countries in the network, and (3) it signs new treaties as new nodes join the network.

4. Fitting a model of network growth

In this section we will use the empirical data to build two models that capture the main aspects of the evolution of the BIT network. To calibrate the parameters used by the models, we define a goodness-of-fit function that measures the ability to reproduce the characteristics of the BIT network. We use these models to shed light on the following questions regarding the nature of BITs: (a) What is the best way to explain the evolution of the BIT network? (b) Is it possible to predict the future behavior of the network? (c) Is the decrease in the rate of new BITs in the period 1995 to 2005 clear evidence of saturation effects or could it have happened because countries were busy dealing with other issues?

We propose a random network growth process, simulate it and compare the outcome with our observations. It is assumed that every country that joined the network remains active and signing BITs forever, and that countries do not reject signing with any other country.[3] We start this growth process by taking a network of one edge connecting two nodes, which represents the first BIT between Germany and Pakistan in 1959. Then, for every $y = \{1960, \ldots, 2005\}$, we add ΔN_y new

[3] These are modeling abstractions that simplify the growth process and represent reality quite accurately. In the real-world, some countries do stop being active, and some countries may have long standing conflicts with others so it may not be possible that a given dyad signs a treaty.

Table 2
Output for Model A and year 2005. Rows represent: fitted estimate for probability r of random treaty, average path length ℓ, diameter, radius, clustering coefficient (CC) and clique number (CN). Row %KS denotes the percentage of instances in the sample for which the degree distribution did not coincide with the empirical one, according to the KS test.

	BITs	F–GPA		PA–GPA		GPA–GPA	
r	–	0.09		0.15		0.22	
ℓ	1.926	1.953	[1.761, 2.121]	1.910	[1.842, 2.101]	1.988	[1.711 2.254]
Diam	4	3.66	[3, 4]	3.8	[3, 4]	3.7	[3, 4]
Radius	2	2	[2, 2]	2	[2, 2]	2	[2, 2]
CC	0.492	0.535	[0.492, 0.584]	0.508	[0.472, 0.545]	0.491	[0.462, 0.522]
CN	14	17.52	[15, 20]	18.133	[16, 21]	23.68	[21, 27]
%KS	–	6%	–	6%	–	8%	–

nodes to the network, where ΔN_y is the actual number of countries that signed their first BIT in year y. This replicates the evolution of the size of the real network. Note that since the network growth is nonlinear in the number of nodes, it is not possible to insert one node per iteration as was done in Refs. [22,7,20]. At the time of incorporation, a node signs one treaty with a country that is already part of the network, chosen using preferential attachment (see Section 3). To complete the new BITs to be signed that year, we add $\Delta T_y = \Delta E_y - \Delta N_y$ edges between the existing nodes, where ΔE_y is the actual number of new BITs in year y. Hence, overall we add the necessary number of arcs to replicate the total number of BITs signed during that year; different processes will generate different topologies but the size of the network will be unaffected.

For each new treaty, we select two random nodes using a distribution $\Pi_y(v)$ and connect them. While we consider the uniform distribution $\Pi_y(v) = 1/N_y$ of Erdős and Rényi as a baseline model, the rest of the distributions are based on the methods described in Section 3.1. When using the fitness of countries, $\Pi_y(v) = \eta(v)/\sum_{w\in N}\eta(w)$ for a fitness vector η. For preferential attachment, $\Pi_y(v) = k_y(v)/2|E_y|$. Finally, when using generalized preferential attachment, $\Pi_y(v) = k_y(v)\eta(v)/\sum_{w\in N_y} k_y(w)\eta(w)$.

To quantify how well the generated graphs fit the real BIT network, we define a goodness-of-fit (GoF) function based on the structural characteristics discussed in Section 2.1 (i.e. average path length, diameter, radius, clustering coefficient, and clique number). More specifically, it is defined as a weighted sum of the relative errors between the estimated means of each parameter and the corresponding empirical value. The weighted sum assigns more importance to the clustering coefficient and to the average path length than to the diameter and clique number because the latter are more sensitive to small changes in the distribution of edges. We generate a sample of 100 observations using the growth process to have an accurate estimation of the means that the GoF function uses. Since the degree distribution is not a scalar, instead of including another term in the GoF function, we check that the resulting degree distribution matches the empirical one. To that extent, we use the *Kolmogorov–Smirnov (KS) test* with a significance level of $\alpha = 0.05$ [23]. Putting it all together, we will conclude that the fit is good whenever the GoF function is small enough and when at least 90% of the instances pass the KS test.

We calibrate our growth process by searching in the space of parameters for those that provide the best fit to BIT_{2005}. This search is done using a local search procedure that we implemented for this purpose. To test the model, we compute 95%-confidence intervals for each structural property for the computed optimal parameters $params_{opt}$. For a successfully calibrated model that agrees with reality, one should have empirical observations that lie within the corresponding confidence intervals.

4.1. Model A

The first model implements the procedure described earlier, for different combinations of distributions $\Pi_y(v)$ for choosing the two endpoints of new interior BITs. The five pairs of criteria that we test are F–F, F–GPA, GPA–GPA, PA–GPA, and PA–PA, where F, PA, and GPA refer to fitness, preferential attachment, and generalized preferential attachment, respectively. In addition, for each new BIT and with probability r, we disregard the chosen criteria and select two countries chosen uniformly at random, similarly to the *ER* model. For cases when a fitness vector η is needed, and since we want to use as little information about the real network as possible, we set $\eta(v)$ to a random value drawn from a distribution $\rho(\eta)$ with parameters $params_{dist}$. The possible distributions ρ that we consider are uniform, normal, and exponential. The number and type of parameters are set according to the type of distribution.

We find the optimal set of parameters for each node selection criterium using our local search procedure. Table 2 presents the actual values and compares them to the means of the topological parameters arising from the simulation with optimal parameters when the fitness distribution is exponential. To the right of the means, we also include the 95%-confidence intervals. Results for F–F and PA–PA are not shown because their fit was poor.

Except for the clique numbers for all criteria, all actual values lie inside the estimated confidence intervals and the percentage of KS tests returning that the empirical and simulated degree distributions were different was small. Hence, Model A generates graphs similar to BIT_{2005} for the three criteria shown in the table. Notice that the clique number equal to 14 in the BIT network is smaller than all intervals for clique numbers. As discussed in Section 2.2, it is possible that some of the treaties that are not signed in the real world due to political reasons, were signed by the growth process since we do not consider that some dyads are not compatible.

Table 3
Output for Model B and year 2005.

	BITs	F–GPA		PA–GPA		GPA–GPA	
r_1, b_1	–	0.2, 0.98		0.2, 0.98		0.2, 0.98	
r_2, b_2	–	0.06, 0.8		0.1, 0.8		0.1, 0.85	
r_3, b_3	–	0.06, 0.5		0.1, 0.5		0.1, 0.5	
ℓ	1.926	1.934	[1.726, 2.095]	1.927	[1.793, 2.102]	1.927	[1.754, 2.338]
Diam	4	3.95	[3, 4]	4.1	[3, 4]	4	[4, 4]
Radius	2	2.02	[2, 2]	2.15	[2, 3]	2	[2, 2]
CC	0.492	0.515	[0.479, 0.524]	0.510	[0.478, 0.530]	0.498	[0.476, 0.529]
CN	14	16.97	[14, 20]	18.5	[17, 21]	24.666	[22, 29]
%KS	–	0%	–	0%	–	0%	–

In addition, we observed that the average path length ℓ was not strongly affected by changes in r. This is not surprising because *ER* graphs induce values of ℓ similar to real-world networks (see Section 2.1) and so adding more "random" arcs should not modify ℓ significantly. On the other hand, the clustering coefficient is decreasing as a function of r because for higher r, edges tend to cluster less around high-degree nodes. Although the best fit for the parameter of the exponential was $\lambda = 1$, the fit is robust to changes in the parameter. Nevertheless, it does depend on the distribution itself; different results were obtained for a uniform, an exponential, and a normal distribution. The best fit was given by an exponential and hence that is the one we selected for our runs.

To test Model A further and determine the accuracy of the network growth process, we stop the simulations at an earlier year and compare the output to the empirical network in the same year. Using the optimal parameters resulting from evaluating the *GoF* function in 2005 and stopping the simulation in years 1975 and 1990, we find that conclusions do not change. As before, the only statistically significant difference between the real and the generated networks is in the clustering coefficient.

Notice that our *GoF* function only takes into account the topology in year 2005. In additional tests, we modified this function to compare the simulated network to the real one in an arbitrary year y. For years 1975 and 1990, we were unable to obtain similar values of the simulated and the real clustering coefficients, concluding that the problem is in fact in the network formation process and not because we are not considering the measures for those years in the functional form of the *GoF*. In theory, one could even define a more precise version that considers the whole evolution of the growth process instead of just the final state but we have not yet experimented with this idea.

4.2. Model B

The model presented in the previous section, although simple, omits an important aspect of the BIT network. As it was discussed in Section 2.1, the BIT network evolution can be divided into phases. Recall that in the first phase treaties were signed mainly between a developed and a developing country, then countries started to sign treaties within the same group, and finally the creation of internal BITs within groups accelerated. By 2005, the number of BITS signed between two developing countries was roughly equal to that between developed and developing countries.

This observation motivates us to consider a more detailed model that improves the fit of the empirical data. We divide time into the three phases $p = \{1, 2, 3\}$ mentioned earlier and allow the parameters to depend on the phase (so now there are three probabilities r_p's of signing a BIT using the *ER* process). Every time a new country joins the BIT network, we randomly assign a development level to it using a Bernoulli process with a probability equal to the proportion of developed and developing countries in 2005. We consider three new parameters b_p that represent the probability a treaty is signed as if the network were bipartite. Indeed, when choosing a dyad randomly as described in Section 4.1, if the development levels coincide, with probability b_p we restart and pick another dyad. We repeat this until a correct dyad is selected.

Table 3 shows the output in 2005. After computing the optimal parameters using our *GoF* function and running the model for different years, we conclude that when using F–GPA criteria, all the measurements lie in the confidence intervals except for the clustering coefficient at year 1990. However, when using the criteria PA–GPA and GPA–GPA, the clique number is not included in the confidence intervals for years 1990 and 2005. Furthermore, the difference in size between actual and simulated clique numbers seems to progressively increase.

Finally, we also assess whether our best model (Model B with node selection criterium F–GPA) can be used to provide accurate predictions. We first find the optimum set of parameters using data up to 2000. Then, we simulate the network up to 2005 and compare the simulation results with the original BITs. The only actual information from the period 2001–2005 used is the set of values of ΔN_y and ΔE_y. We observe that all the original measures of BITs lie inside the computed confidence intervals.

4.3. Quasi-cliques

Although the criterium F–GPA provided the most accurate results, this was driven in part by the relatively bad fit of clique numbers with the other criteria. As discussed earlier, just a few missing edges in the original graph can introduce a

Table 4
Density (γ) and quasi-cliques number (QCN) for the original network and the best instance for each selection criterium, as a function of the maximum number of missing edges n allowed in the quasi-clique.

n	BITs		F–GPA		PA–GPA		GPA–GPA	
	γ	QCN	γ	QCN	γ	QCN	γ	QCN
0	1	14	1	16	1	17	1	24
5	0.967	18	0.970	19	0.976	21	0.983	25
10	0.941	19	0.947	20	0.960	23	0.971	27
20	0.913	22	0.920	23	0.938	26	0.954	30
50	0.857	27	0.857	27	0.885	30	0.910	34

significant difference in the empirical clique numbers. This motivates us to look at quasi-clique numbers because they are a more robust measure of the cohesiveness of a group of countries.

We look at quasi-cliques with a perspective of bicriteria optimization where the two objectives are the size of the group and its density, and determine the Pareto-curve. This can be done by computing the values of $f(n)$, defined as the size of the maximum quasi-clique in a graph G that has at most n edges missing. Since this problem is strongly NP-hard, we use an integer programming formulation and a cut generation procedure to compute these values (see Ref. [19] for details). The high computational cost makes it prohibitive to compute values for the whole sample. Hence, we just compute f for the best instance in the sample, for each node selection criterium. To select the best instance among those generated with the optimum set of parameters, we look for the simulated graph minimizing the *GoF* function among those whose properties lie inside the confidence intervals.

Table 4, which shows the quasi-clique numbers $f(n)$ for $n = 5, 10, 20, 50$ and the different node selection criteria, confirms the conclusions we drew when we analyzed Model B and discards the alternative explanation that the difference was due to a few extra edges. Indeed, while the behavior of the real network and F–GPA are almost the same when allowing n edges to be missing, we can observe that the maximum quasi-cliques for GPA–GPA and PA–GPA are larger. Hence, the ultimate conclusion is that the process that generates a graph most similar to the actual BIT network is F–GPA.

4.4. Endogenizing the edge creation process

We have built a model that reproduces the characteristics of the BIT network accurately. However, this model is based on the actual values of ΔN_y and ΔE_y. One interesting question that arises naturally is whether we can replicate the growth process, and particularly its nonlinear behavior, without using the time series of new BITs per year. In this section we propose a process that endogenizes arc creation and sheds light on the issue of saturation. Indeed, since the number of new treaties it generates in the period 1995–2005 tends to decline, our conclusion is that countries are running out of candidates with which to sign BITs. This should be contrasted with the alternative explanation that some external factor such as wars or economic crises forced politicians and diplomats to shift their interest from considering new countries with which to sign BITs to other endeavors.

To enrich the model as described, we adapted our best model (Model B with criterium F–GPA) so it considers signing ΔT_y new internal treaties in year y. Instead of coming from the actual data as before, ΔT_y is now defined to be a small fraction ϵ of the possible number of treaties that can be signed, specifically $\Delta T_y = \epsilon |N_y|(|N_y| - 1)/2$. In every iteration a dyad is selected for consideration. If a BIT has already been signed between them or if the two countries do not have the correct development levels, the opportunity is lost and a new dyad is considered. According to this process, the number of new treaties in year y can be any number between 0 and ΔE_y, instead of a fixed number as before. We highlight that the number of new nodes per year ΔN_y still comes from historical data because we needed a source of nonlinearity to replicate the empirical data.

The resulting number of BITs is plotted in Fig. 7 alongside with the real time series. Both curves are very close to each other. Also, in the last years of the simulation, the model attempts to sign around 600 BITs out of which less than 100 are successfully signed. The reason for this is that most candidate treaties are either already signed or are not feasible. This is strong evidence supporting the network saturation hypothesis. At the same time, the decrease in the coefficient of variation and the change in the shape of the degree distribution are probably consequences of this saturation.

5. Conclusions

We have presented a detailed study of the BIT network using the framework of social networks. The most important features that make this network different from others studied in the literature is that our network is small, dense, and shows signs of saturation. The techniques we have put forward could also be relevant to other small networks with these properties.

Our empirical observations on the BIT network in the period 1959–2005 have demonstrated that it shows the small-world property at all times and the average path length is almost constant. On the other hand, the clustering coefficient and the degree distribution vary over time. While the clustering coefficient is lower than expected for an *ER* graph with the same number of nodes and a similar number of edges at an early stage, from 1985 onward we have seen that it is considerably higher, a property which is shared by most real-world networks. Although the node degree distribution is skewed right and

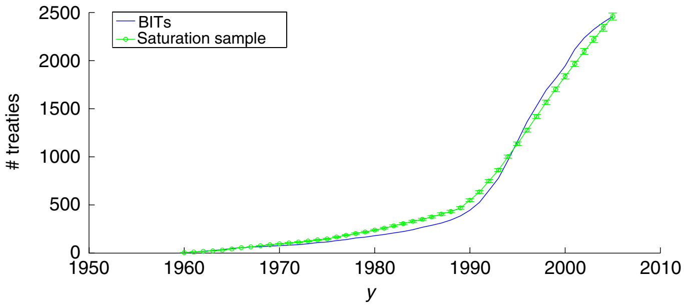

Fig. 7. Comparison of the number of treaties signed in the BIT network, to the number of treaties signed by the modified Model B, using $\Delta T_y = 0.038|N_y|(|N_y| - 1)/2$.

has a long tail across the whole period, it changes from approximately power-law in 1975 and 1990 to exponential in 2000 and 2005. We have provided pieces of evidence that support that this change is due to signs of saturation.

While most previous work on the proliferation of bilateral investment treaties has focused on the correlation between BITs and FDI, our analysis suggests that the signing of treaties can be explained by a generalized preferential attachment process. Our model is based on one proposed by Barabási et al. to detect signs of preferential attachment at both new node and internal edge levels [7]. The new modeling elements include considering three distinct phases and classifying nodes as developed or developing to force the network to be partially bipartite. These elements have been found to be a fundamental driver of the BITs evolution since otherwise the model provides clique numbers that do not match empirical observations. Finally, we have also explored a procedure that makes the edge creation process endogenous. This was used to provide evidence supporting the existence of saturation because countries have already signed most BITs that interest them.

This work is the first to propose the use of quasi-cliques as an instrument to evaluate differences in the topology of random graphs. Quasi-cliques have provided us with concrete evidence that different processes may generate graphs with very similar global structural properties, such as average path length, clustering coefficient and degree distribution, and yet can still have notorious differences in their local structure.

The models introduced in this work are only a first step towards more complete models of the BIT network. Below, we list some elements left for future work. (a) The fitness value of a country could vary over time to reflect changing policies and political situations. (b) Fitness values for each country were drawn from a probability distribution. Instead, one could try to explain the fitness value using relevant explanatory variables. (c) Incorporate prohibited dyads, which certainly exist in the real world. (d) Multilateral Investment Treaties can also be considered using hypergraphs. (e) We have neglected to examine whether BITs were ratified. Adding this data could be used to gauge the interest that both countries have in the BIT.

Acknowledgements

We thank Pablo Pinto for motivating us to pursue this research and for all his knowledgeable advice, and Matías Graña and Pablo Jacovkis for valuable suggestions that improved this work. The first and second authors acknowledge support from ANPCyT PICT-2007-00518 and 00533, Argentina. The second author was also partially supported by UBACyT Grants X069 and X606, Argentina. Finally, the third author was partially supported by the Columbia University Center for International Business Education and Research.

References

[1] T. Büthe, H.V. Milner, The politics of foreign direct investment into developing countries: increasing FDI through trade agreements? American Journal of Political Science 52 (4) (2008) 741–762.
[2] Z. Elkins, A.T. Guzman, B.A. Simmons, Competing for capital: the diffusion of bilateral investment treaties, 1960–2000, International Organization 60 (2006) 811–846.
[3] E. Neumayer, L. Spess, Do bilateral investment treaties increase foreign direct investment to developing countries? World Development 33 (10) (2005) 1567–1585.
[4] S. Rose-Ackerman, J. Tobin, Foreign direct investment and the business environment in developing countries: the impact of bilateral investment treaties, Yale Law & Economics Research Paper, No. 293, 2005.
[5] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the internet topology, ACM SIGCOMM Computer Communication Review 29 (4) (1999) 251–262.
[6] R. Albert, H. Jeong, A.-L. Barabási, The diameter of the World Wide Web, Nature 401 (1999) 130–131.
[7] A.-L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A 311 (3) (2002) 590–614.
[8] H. Jeong, S.P. Mason, A.-L. Barabási, Z. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.
[9] P. Erdős, A. Rényi, On random graphs, Publicationes Mathematicae 6 (1959) 290–297.
[10] R. Albert, A.-L. Barabási, Topology of evolving networks: local events and universality, Physical Review Letters 85 (24) (2000) 5234–5237.
[11] M.A. Serrano, M. Boguñá, Topology of the world trade web, Physical Review E 68 (2003) 015101.
[12] D. Garlaschelli, M.I. Loffredo, Structure and evolution of the world trade network, Physica A 355 (1) (2005) 138–144.
[13] D.J. Fenn, O. Suleman, J. Efstathiou, N.F. Johnson, How does Europe make its mind up? Connections, cliques, and compatibility between countries in the Eurovision Song Contest, Physica A 360 (2) (2006) 576–598.

[14] UNCTAD: United Nations Conference on Trade and Development, Bilateral Investment Treaties 1959–1999, 2000.
[15] UNCTAD: United Nations Conference on Trade and Development, Investment Instruments Online — Bilateral Investment Treaties, 2008.
[16] D. West, Introduction to Graph Theory, 2nd ed., Prentice Hall, 2000.
[17] D.J. Watts, Small Worlds, Princeton University Press, Princeton, 1999.
[18] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences), Cambridge University Press, 1994.
[19] F. Bonomo, J. Marenco, D. Saban, N.E. Stier-Moses, A polyhedral study of the maximum edge subgraph problem, Electronic Notes in Discrete Mathematics 35 (2009) 197–202.
[20] G. Bianconi, A.-L. Barabási, Competition and multiscaling in evolving networks, Europhysics Letters 54 (4) (2001) 436–442.
[21] L.A. Amaral, A. Scala, M. Barthelemy, H.E. Stanley, Classes of small-world networks, Proceedings of the National Academy of Sciences 97 (21) (2000) 11149–11152.
[22] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[23] D. Boes, F. Graybill, A. Mood, Introduction to the Theory of Statistics, McGraw-Hill, New York, 1974.