

# Topology and evolution of the network of western classical music composers

Doheum Park<sup>1</sup>, Arram Bae<sup>1</sup>, Maximilian Schich<sup>2</sup> and Juyong Park<sup>1\*</sup>

\*Correspondence:

juyongp@kaist.ac.kr

<sup>1</sup>Graduate School of Culture Technology and BK21 Plus Postgraduate Organization for Content Science, Korea Advanced Institute of Science & Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea  
Full list of author information is available at the end of the article

## Abstract

The expanding availability of high-quality, large-scale data from the realm of culture and the arts promises novel opportunities for understanding and harnessing the dynamics of the creation, collaboration, and dissemination processes - fundamentally network phenomena - of artistic works and styles. To this end, in this paper we explore the complex network of western classical composers constructed from a comprehensive CD (Compact Disc) recordings data that represent the centuries-old musical tradition using modern data analysis and modeling techniques. We start with the fundamental properties of the network such as the degree distribution and various centralities, and find how they correlate with composer attributes such as artistic styles and active periods, indicating their significance in the formation and evolution of the network. We also investigate the growth dynamics of the network, identifying superlinear preferential attachment as a major growth mechanism that implies a future of the musical landscape where an increasing concentration of recordings onto highly-recorded composers coexists with the diversity represented by the growth in the sheer number of recorded composers. Our work shows how the network framework married with data can be utilized to advance our understanding of the underlying principles of complexities in cultural systems.

**Keywords:** complex network; classical music; network topology; network evolution

## 1 Introduction

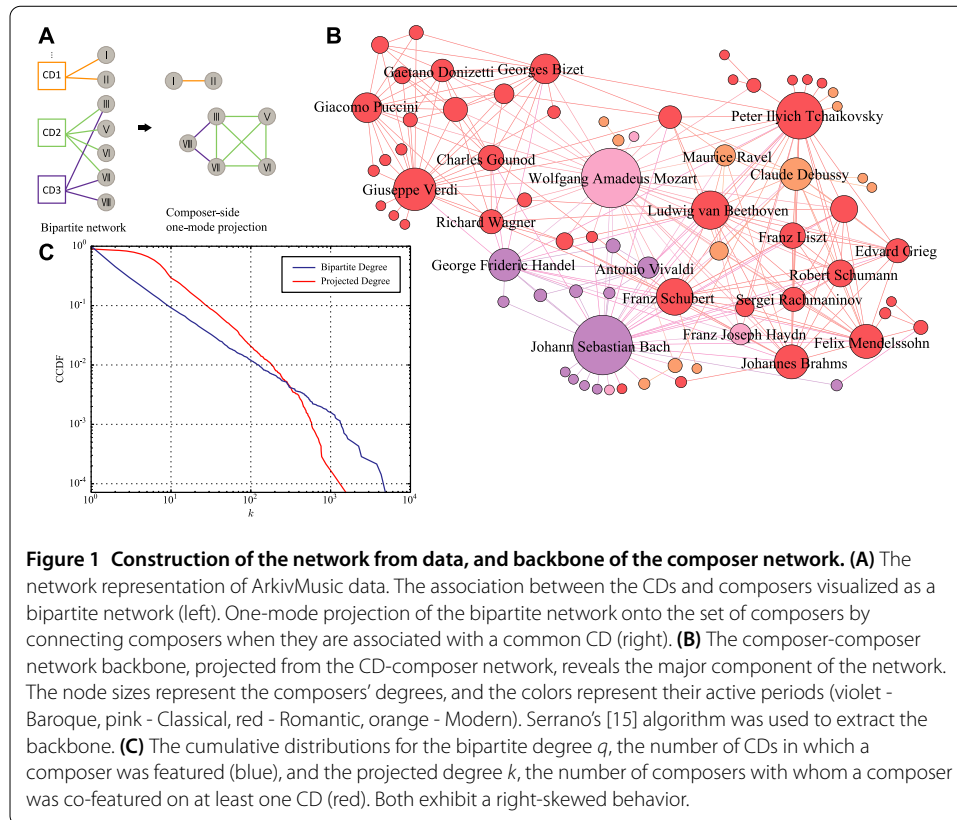
Networks have been used extensively in recent years for characterizing and modeling intricate patterns found in various social, technological, and biological complex systems originating from the functional and dynamical interdependence between their components. The methodology is expanding its horizon, being eagerly adopted in new fields such as culture for exploring novel answers to new and long-standing issues [1–4]. In broad terms, the potential of the network framework for understanding culture originates from the observation that the creation and transmission of cultural products are essentially network phenomena. Networks, therefore, may lead to a new fundamental understanding of the complex nature of culture.

There are several interesting questions one can explore regarding such ‘network in culture’ - e.g., How does it evolve and change over time? Who are the most prominent or popular artists, and how do we measure their importance? How do the different styles combine to produce cultural products? How does information flow over network ties? [5]

- the answers to which would contribute to deepening our understanding of culture and the arts.

There have been several notable works that highlight the importance of networks on understanding various issues in culture. Suárez, Sancho and de la Rosa [6] analyzed the linkage patterns from the data of 11,443 artworks from Spain and Latin America with respect to genre and theme, identifying religious theme to be a dominant factor connecting the paintings. Gleiser and Danon [7] studied the topology and the community structure of the collaboration network of Jazz musicians, uncovering the presence of communities based on the locations of the bands correlated with the racial segregation between the musicians. Park *et al.* [8] compared two networks of contemporary popular musicians - one representing the collaboration history of the musicians, and the other representing musical similarities as judged by human experts - finding that the significant topological differences are related closely to the nature of the connections. Salganik *et al.* [9] studied the role of social influence and success of cultural products. They found that social influence has a significant effect, elevating the inequality and the unpredictability of success. Uzzi and Spiro [10] investigated the small-world property of the network of the creators of original Broadway musicals that reflect the level of their cohesion and its impact on the success of the musicals. They found that cohesion had a positive impact up to some point, but began to impede the creativity when it becomes too large. The characteristics of a system and growth or evolutionary dynamics are deeply intertwined [11]. This is likely to be the case for cultural system as well, and grasping the underlying growth principles may need a better understanding of their nature. This line of thinking is behind the work of Jeong *et al.* [12] who applied network evolution models to the movie-actor network. Recently in the field of music, Serrà *et al.* [13] constructed a time-varying network of pitch transitions in contemporary western popular music (from 1955 to 2010) and analyzed the degree distributions of pitches, pointing out the limited use of various pitch transitions.

In this paper we analyze the co-occurrence network of western classical composers constructed from the comprehensive data set of CD recordings. Using the network framework, we try to shed light on the following specific questions in order: Who are the prominent composers? What are the driving force behind composers being co-featured on a common CD? Can we characterize the temporal growth of the network and the composers? We start by measuring the fundamental network properties that give us a bird's-eye view of the general features of the network, including the size, clustering coefficient, assortativity, degree distribution, centralities, and community structure. The centralities identify influential composers with varying artistic styles, while the positive assortativity shows that period designations and artistic styles are the main driving force behind co-featured composers. This is explored further using the community structure of the composers. We characterize the growth of the composer-CD network by way of the temporal evolution of the bipartite degree distribution, and how it can be mapped to a superlinear preferential attachment model. It also allows us to forecast the future of the landscape of the composer network, where the growths of the degrees of prominent composers are accelerating and accordingly the recordings are becoming increasingly concentrated on those already well established, an effect similar to the so-called urban-scaling laws, where existing cultural centers win out over new centers [4, 14].



## 2 Data and network construction

We utilized two online data resources: ArkivMusic (<http://www.arkivmusic.com>), an online music retailer, and All Music Guide (<http://www.allmusic.com>), a comprehensive musical information provider. As of 2013, ArkivMusic lists more than 96,000 classical music CDs and its title, release date, and the composers and performers of the music. For this work we specifically use the title, release date and the composers. As we show in Figure 1(A), the data can be represented as a *bipartite network* composed of two node classes - CDs and composers - where an edge is drawn when a composer's pieces were recorded on the CD. A *one-mode projection* onto the composer class results in a network solely of composers in which two nodes are connected if they have been co-featured on a CD (Figure 1(A)). Figure 1(B) shows the so-called network backbone (a scaled-down representation [15]). Compilation CDs (a repacked collection of previous released recordings) and the ones without release dates were weeded out from the data, resulting in 63,679 CDs and 13,981 composers for analysis.

## 3 Topology of the composer network

First we examine the fundamental properties of the composer network to understand its general features. We then analyze the relationship between the characteristics of connected composers to study how they affect the chances of connection, i.e. becoming featured on a common recording. We investigate the centralities, the assortativity, and the community structure of the composer-composer network to answer the question of who are the prominent composers, and what drives the formation of connection patterns.

### 3.1 Fundamental network properties reveal the small-world property, a high clustering, and wide differences in composer prominence

Many networks exhibit the so-called ‘small-world’ property [16], often expressed using ‘six degree of separation’ in common parlance. Mathematically defined as the mean geodesic distance (the length of the shortest path) between nodes increasing sublinearly (typically logarithmically) as a function of the size of the network, it is frequently represented by a small mean geodesic distance (a geodesic is the shortest path between two nodes. If they are directly connected, the geodesic distance is 1). Technically speaking the two definitions are unidentical [17], although for the purpose of our paper the distinction is not critical. The mean geodesic distance and the diameter (the longest geodesic) in the largest component (set of connected nodes) are 13.1 and 20 for the CD-composer network, and 3.5 and 20 for the composer-composer network, respectively.

The *clustering coefficient*  $C$  is the probability that two nodes connected to a common neighbor are themselves neighbors. It is a measure of the abundance of triangles in the network, and a high value is a hallmark of social networks [18]. In our composer-composer network we have  $C = 0.648$  compared with the random expectation  $\bar{C}_{\text{random}} = 0.001$ , meaning that our network is indeed highly clustered. Such a clustered behavior is deeply related to patterns of mixing and community structure, which we later discuss in detail.

We now investigate the nodes’ *degrees* in the network. A node’s degree, the number of its neighbors, is a type of its *centrality*, a class of quantities that signify a node’s importance or prominence determined from the network. The degree is the basic and most intuitive centrality, and is the principal factor behind many network properties [17]. As such, it can be useful in understanding the different roles or positions of the composers in our network. Specifically, in our study we consider two types of composer degrees: The bipartite degree  $q$  in the CD-composer network, and the projected degree  $k$  in the composer-composer network. While closely related, they carry slightly different yet interesting probable interpretations: The bipartite degree would mainly represent a composer’s popularity, while the projected degree would mainly represent a composer’s versatility or compatibility with others, likely in terms of musical styles. The nodes’ degrees as a whole can be characterized by plotting the degree distributions  $p(q)$  and  $p(k)$ , or the cumulative distributions  $P(q) = \sum_{q'=q}^{\infty} p(q')$  and  $P(k) = \sum_{k'=k}^{\infty} p(k')$  plotted in Figure 1(B) on a log-log scale. We note that both distributions exhibit heavy tails, i.e. a few composers take up the lion’s share of recordings and composer-composer associations: In the CD-composer network, Wolfgang Amadeus Mozart (1756-1791) has the highest degree  $q = 4,851$ , approximately 505 times the mean  $\bar{q} = 9.6$ , followed by Johann Sebastian Bach (1685-1750) with  $k = 4,292$ . In the composer-composer network JS Bach has the highest degree,  $k = 1,551$ , approximately 103 times that of average degree  $\bar{k} = 15.1$ , followed by WA Mozart with  $k = 1,086$ . For reference, in Table 1 we show the names of composers with the twenty highest  $q$  and  $k$  values. The heavy-tailed degree distribution in Figure 1(B) tells us that there is a significant variation in the composers’ prominence and popularity in the landscape of classical music, which provides us with an important clue for understanding its nature and how it would potentially evolve in time, which we discuss in a later chapter.

### 3.2 Composer centralities reveal the relationship between the network and composer characteristics

Next, we study further types of composer centralities in the composer-composer network that capture various ‘flavors’ of a node’s importance in a network.

**Table 1 Top 20 composers for degree, eigenvector, and betweenness centralities**

Rank	Bipartite degree ( $g$ )		Projected degree ( $k$ )		Eigenvector centrality		Betweenness centrality	
	Name	Period	Name	Period	Name	Period	Name	Period
1	WA Mozart	C	JS Bach	B	JS Bach	B	JS Bach	B
2	JS Bach	B	WA Mozart	C	WA Mozart	C	WA Mozart	C
3	Beethoven	R	Handel	B	Handel	B	Handel	B
4	Brahms	R	Brahms	R	Brahms	R	Piazzolla	M
5	Schubert	R	Mendelssohn	R	Mendelssohn	R	Brahms	R
6	Verdi	R	Debussy	M	Schubert	R	Gershwin	M
7	Tchaikovsky	R	Schubert	R	Debussy	M	Debussy	M
8	R Schumann	R	Beethoven	R	Beethoven	R	Mendelssohn	R
9	Handel	B	Saint-Saëns	R	Saint-Saëns	R	Schubert	R
10	Wagner	R	Tchaikovsky	R	Tchaikovsky	R	Beethoven	R
11	Chopin	R	Ravel	M	Ravel	M	Villa-Lobos	M
12	Haydn	C	Gershwin	M	Fauré	R	Ravel	M
13	Liszt	R	R Schumann	R	R Schumann	R	Tchaikovsky	R
14	Mendelssohn	R	Fauré	R	Liszt	R	Copland	M
15	Debussy	M	Liszt	R	Chopin	R	Saint-Saëns	R
16	Puccini	R	Vivaldi	B	Vivaldi	B	Vivaldi	B
17	Vivaldi	B	Piazzolla	M	Rossini	R	Stravinsky	M
18	Dvořák	R	Rossini	R	Rachmaninoff	M	Britten	M
19	Ravel	M	Chopin	R	Haydn	C	Hindemith	M
20	R Strauss	R	Verdi	R	Gershwin	M	Bernstein	M

Each centrality can be interpreted as representing distinct composer characteristics: The bipartite degree represents a composer's popularity; the projected degree represents a composer's compatibility with others; the eigenvector centrality is a generalization of the degree that considers the quality of connections (e.g. whether a composer tends to be paired with other prominent composers); the betweenness centrality measures how often a composer acts as an intermediary between two composers. Periods are abbreviated: Baroque (B), Classical (C), Romantic (R), and Modern (M).

Besides the degree that we have already seen, the *Eigenvector Centrality* and the *Betweenness Centrality* (also called the *Freeman Centrality* after its inventor) [19, 20] are widely used. The eigenvector centrality is a generalization of the degree that considers the 'quality' of a connection: Being a neighbor to a central node in turn raises one's own eigenvector centrality. The name comes from its mathematical definition as the components of the leading eigenvector of the adjacency matrix. The betweenness centrality measures how often a node sits on the geodesic between two nodes, acting as an intermediary (e.g., in communication). One benefit of investigating different centralities is that while the centralities are often correlated, significant disagreements can point to unusual aspects of the network that in turn can lead to a deeper understanding of it. The highest-ranked composers in each centrality are given in Table 1. The lists do appear correlated, with Spearman Rank Correlations (SPR) equal to  $0.753 \pm 0.002$  between degree  $k$  and eigenvector centrality, and  $0.784 \pm 0.003$  between degree  $k$  and betweenness. But composers labeled 'M' for Modern (the composer's *period*, to be discussed later) are ranked significantly higher in betweenness than in other centralities. It turns out that Modern composers form a tight-knit group with many connections between them, elevating the betweenness of prominent Modern composers such as Aaron Copland (1900-1990) and Leonard Bernstein (1918-1990) although their degree is significantly lower than those from other periods. In order to understand the implications of this type of relationship between a node attribute and network topology, next we review the common period designation in western classical music and analyze them further.

### 3.3 Common artistic style and period designations in western classical music

As one of the oldest art forms, music has a rich history of academic research and investigation [21–24]. While not every expert would agree on one single scheme, it is common to break down the evolution of western classical music into the following several stages by distinguishable styles [25, 26] (we follow the convention employed by All Music Guide, all years are approximate and in CE):

1. *Medieval* (500-1400). The period when primeval shape of musical notation appeared, along with advances in tonal material, texture, and rhythm. Polyphony took shape in terms of tonal material [21]. Notable composers include Guillaume de Machaut (1300-1377) and Francesco Landini (1325-1397).
2. *Renaissance* (1401-1600). The period of modes and rich textures in four or more parts blending strands in the musical texture, harmony, and progression of chords [27]. Notable composers include Thomas Tallis (1505-1585), William Byrd (1540-1623), and John Dowland (1563-1626).
3. *Baroque* (1601-1750). The period distinguished by the creation of tonality. During this period, composers used elaborate musical ornamentation and made changes in musical notation. Baroque music became more complex and expanded the range of instrumental performance [23]. Notable composers include Henry Purcell (1659-1695), Antonio Vivaldi (1678-1741), Johann Sebastian Bach (1685-1750), and George Frideric Handel (1685-1759).
4. *Classical* (1730-1820). The period characterized by a lighter, clearer texture than Baroque. Variety and contrast within a piece became more pronounced than before, and melodies tended to be shorter, with clear-cut phrases and clearly marked cadences [28]. Notable composers include Wolfgang Amadeus Mozart (1756-1791) and Franz Joseph Haydn (1732-1809).
5. *Romantic* (1815-1910). The period when music was closely related with romanticism, the artistic and literary movement in Europe [29]. Romantic music is characterized by freedom of form, emotions, individuality, dynamic changes and nationalism. Notable composers include Ludwig van Beethoven (1770-1827), Franz Schubert (1797-1828), Frédéric Chopin (1810-1849), Robert Schumann (1810-1856), Franz Liszt (1811-1886), and Pyotr Ilyich Tchaikovsky (1840-1893).
6. *Modern* (1900-current). The period characterized by musical innovations in organizing and approaching harmonic, melodic, sonic, and rhythmic aspects leading to many novel styles including expressionism, abstractionism, neoclassicism, futurism, *etc.* [30]. The rise of American classical music was also significant. Notable composers from this period include Claude Debussy (1862-1918), Maurice Ravel (1875-1937), Sergei Rachmaninoff (1873-1943), Igor Stravinsky (1882-1971), George Gershwin (1898-1937) and Leonard Bernstein (1918-1990).

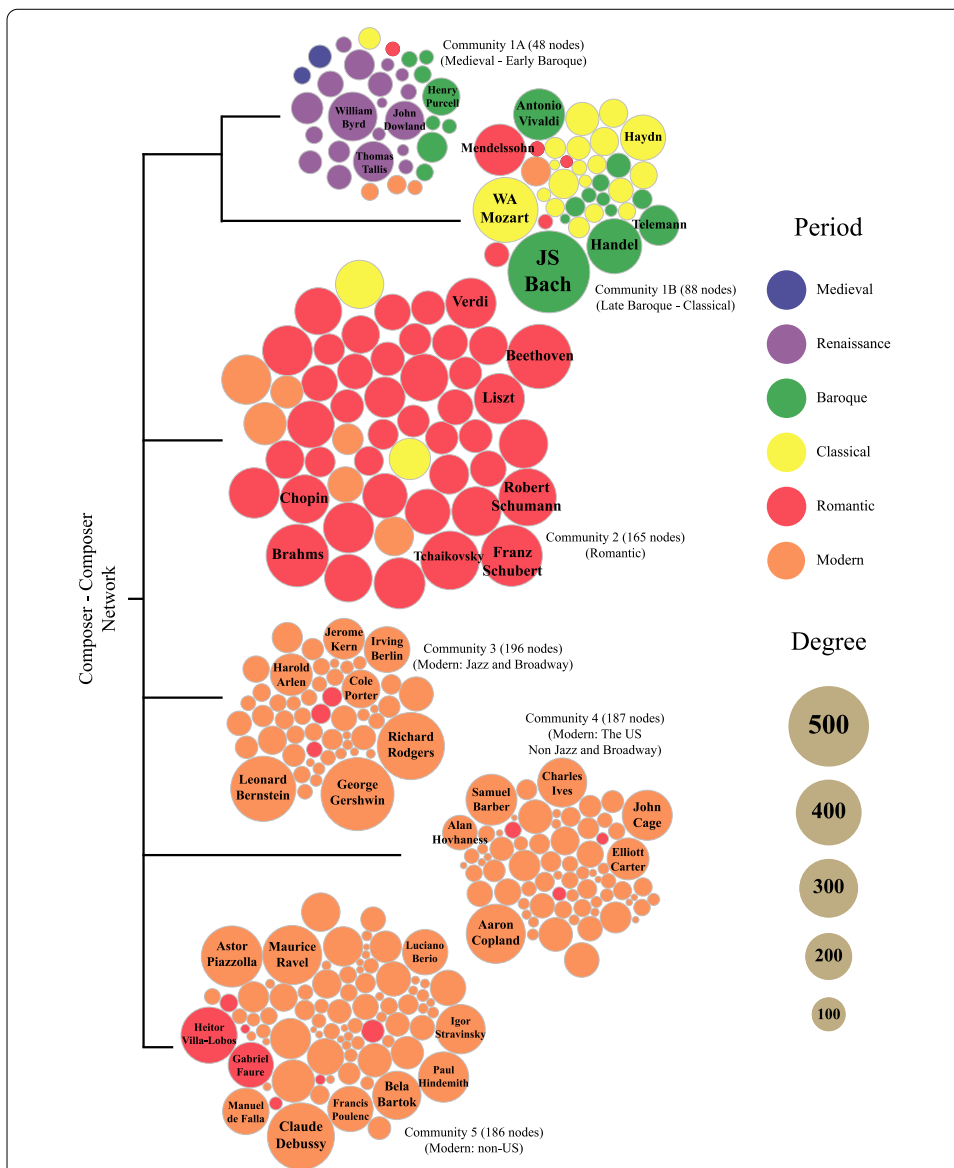
The composer metadata (period and active years) were available for 878 composers, leaving us with 13,667 edges between those in the composer-composer network. While accounting for 6.3% of the entire composer group, these are still the most prominent and significant ones who would be of primary interest; the average bipartite degree for this group is  $\bar{q} = 64.8$ , nearly twenty times larger than the remainder for which  $\bar{q} = 3.5$ .

### 3.4 Assortativity and community structures reveal artistic styles and periods as the main factor behind connections between composers

The relationship between node characteristics and network topology can be quantified by the assortativity coefficient for discrete node characteristics [31] given by  $r \equiv \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$ , where  $\mathbf{e} = \{e_{ij}\}$  is a matrix whose element  $e_{ij}$  is the fraction of edges in a network that connect a node of type  $i$  to one of type  $j$ , and  $a_i$  and  $b_i$  are the fraction of each type of end of an edge that is attached to nodes of type  $i$ . For the composers' periods we have  $r = 0.257 \pm 0.005$ , meaning that composers belonging to a common period tend to be connected preferentially to one another. The Pearson Correlation Coefficient (PCC) between connected composers' active years (the middle point between their birth and death years) is even higher, with  $0.451 \pm 0.009$ .

The assortative mixing we see here is intimately related to the existence of *communities* or *modules* in a network. A community is commonly defined as a group of nodes of a network in which connections are denser than randomly expected or to the rest of the network. Algorithms for detecting communities have seen significant developments in recent years [32–36]. We used the Louvain algorithm of Blondel *et al.* [37] on our reduced network, among many excellent choices. We examined the five largest communities that accounted for 99% of the 878 nodes with known periods. We split Community 1 further into two (named 1A and 1B) since unlike other communities, Community 1 was ostensibly a mixture of many periods. This leaves us with the six sizable communities shown in Figure 2. We find that the communities roughly correspond to the periods introduced earlier; i.e. Community 1A to Renaissance and early Baroque, 1B to late Baroque and Classical, 2 to Romantic, and 3, 4, 5 to Modern (differentiated further among themselves as presented below). We note that, while in terms of sheer number 1B contains more Modern composers, they are rather insignificant compared with others, mainly late Baroque and Classical (the mean degree of Modern composers in the Community is 19.9, while it is 77.8 for the rest). The notable composers in each community are

- Community 1A: William Byrd (1540-1623, Renaissance) and Henry Purcell (1659-1695, Baroque)
- Community 1B: Antonio Vivaldi (1678-1741, Baroque), Johann Sebastian Bach (1685-1750, Baroque), George Frideric Handel (1685-1759, Baroque) from the Baroque period, and Wolfgang Amadeus Mozart (1756-1791, Classical), and Franz Joseph Haydn (1732-1809, Classical) from the Classical period.
- Community 2: Ludwig van Beethoven (1770-1827) and Franz Schubert (1797-1828) who are considered transitional between Classical and Romantic; Robert Schumann (1810-1856, Romantic), Frédéric Chopin (1810-1849, Romantic), Franz Liszt (1811-1886, Romantic), Johannes Brahms (1833-1897, Romantic), and Pyotr Ilyich Tchaikovsky (1840-1893, Romantic) from Romantic.
- Community 3: A US-centric Modern community, with two highest-degree Modern composers being George Gershwin (1898-1937, Modern) of *Rhapsody in Blue* and Leonard Bernstein (1918-1990, Modern) of *West Side Story*. Scott Joplin (1867-1917, Modern) and Billy Strayhorn (1915-1967, Modern), both prominent Jazz composers, and Richard Rodgers (1902-1979, Modern) and Irving Berlin (1888-1989, Modern), both Broadway composers, are also included.
- Community 4: Another US-centric community. Including the likes of Charles Ives (1874-1954, Modern) of *The Unanswered Question*, Aaron Copland (1900-1990,



**Figure 2 The community structure of the composer-composer network.** The five largest composer communities identified using the Louvain method [37] (covering 6.2% of the composers who account for 60.1% of degrees) correspond well to the established period definitions in classical music history and musicological literatures. For each community we list its size and the major represented period(s). Communities 1A and 1B are subcommunities of Community 1, and correspond to the earlier and the later periods of Medieval and Classical. Community 2 represents the Romantic period, and Communities 3, 4, and 5 represent Modern composers. A closer inspection further differentiates the Modern communities: Community 3 represents US Jazz and Broadway composers, while Community 4 represents the other Modern US composers. Community 5 represents the non-US branch of Modern music.

Modern) of *Appalachian Spring*, Samuel Barber (1910-1981, Modern) of *Adagio for Strings*, and John Cage (1912-1992) of 4'33'', this can be said to represent the 20th-century American vernacular style of classical music [38]. More contemporary US composers, Terry Riley (1935-current, Modern), Steve Reich (1936-current, Modern) and Philip Glass (1937-current, Modern), are also in this module.



- Community 5: Composed of mainly Modern (89.3%) and Romantic (10.2%) composers, it includes transitional figures such as Gabriel Fauré (1845-1924, Romantic), Claude Debussy (1862-1918, Modern), and Maurice Ravel (1875-1937, Modern). In contrast with Community 4, this community represents the non-US branch of modern music, including Arnold Schoenberg (1874-1951, Modern, Austria), Manuel de Falla (1876-1946, Modern, Spain), Béla Bartók (1881-1945, Modern, Hungary), Igor Stravinsky (1882-1971, Modern, Russia), Heitor Villa-Lobos (1887-1959, Modern, Brazil), Paul Hindemith (1895-1963, Modern, Germany), Francis Poulenc (1899-1963, Modern, France), Ástor Piazzolla (1921-1992, Modern, Argentina), and Luciano Berio (1925-2003, Modern, Italy).

That the communities appear to correspond well to periods indicate the existence of correlation between the two partitions. The overlap can be quantified, for instance, via the normalized mutual information measure proposed by Danon *et al.* [36, 39]. It is given as  $I_{\text{norm}}(\mathcal{X}, \mathcal{Y}) = \frac{2I(X, Y)}{H(X) + H(Y)}$ , where  $\mathcal{X}$  and  $\mathcal{Y}$  are the partitions of the nodes by community detection and period designations,  $I(X, Y)$  is the mutual information, and  $H(X)$  and  $H(Y)$  are the Shannon entropies of  $X$  and  $Y$ . Widely used in tests of community detection algorithms, the normalized mutual information equals 1 if the partitions are identical and 0 if the partitions are independent. In other words, in our network of composers, the normalized mutual information is 1 if the members in each community are completely identical in periods, and 0 if they are uncorrelated. The normalized mutual information of our result using Louvain method is 0.325, indicating a significant level of overlap, as was hypothesized.

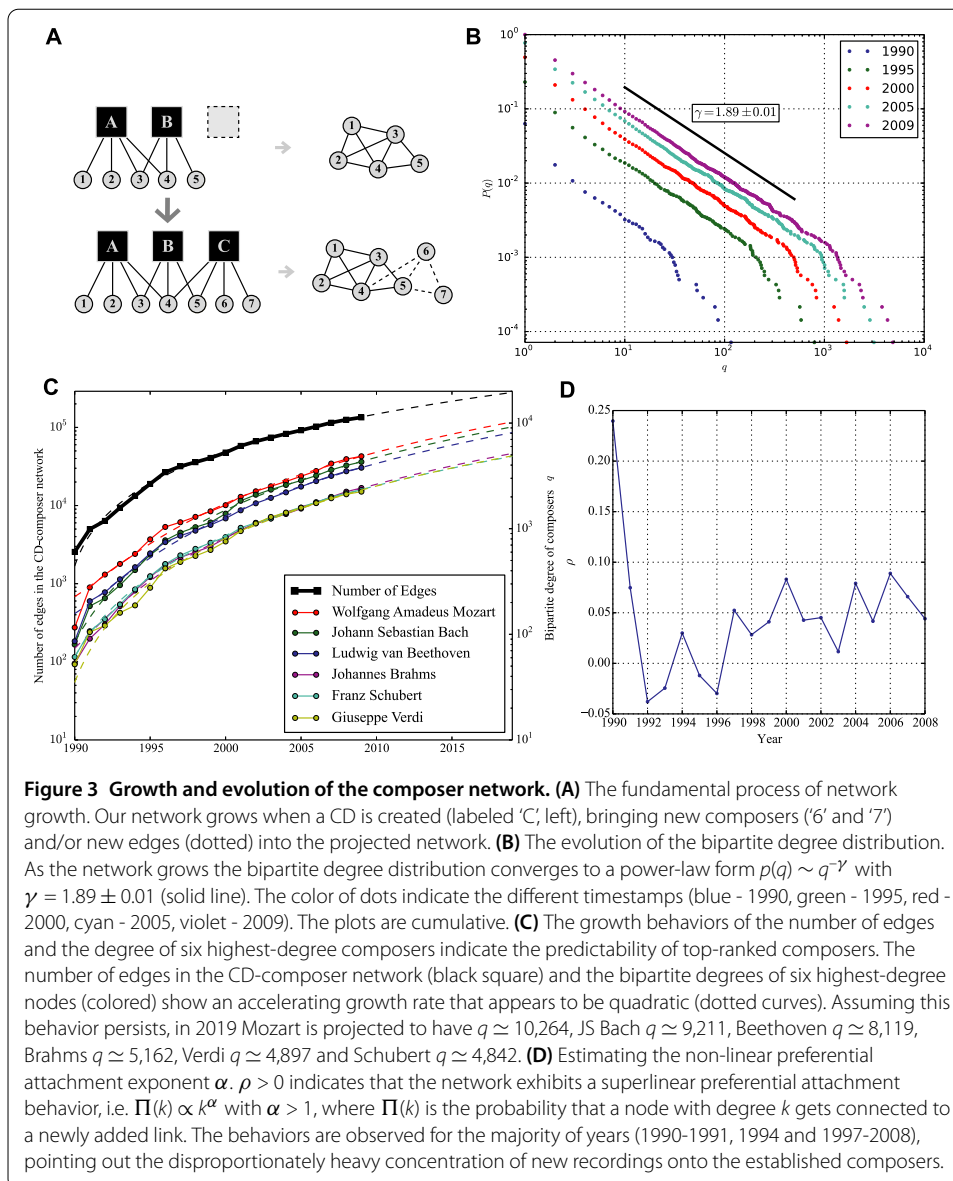
Our investigation of the centralities, assortativity, and community structure shows how large-scale data built primarily for commercial purposes can yield a coherent and useful picture of the landscape of western classical music. This demonstrates that the quantitative analysis of a large collection of cultural artifacts such as CDs can indeed yield meaningful results, validated by agreements with qualitative musicology. A deeper understanding of the network of artists based on commercial cultural products may help in devising new ways of approaching the market which can in turn result in larger, more elaborate commercial data that can further help advance our understanding of the subject.

#### 4 Growth and evolution of the composer network

The classical musical sphere is constantly evolving, with new composers entering the scene, and old composers gaining further prominence or fading out in popularity. As a consequence it is interesting to understand the dynamics of the evolution of popularity or success [40].

##### 4.1 Network growth process hints at the uneven growth of oldboys and newbies

The network growth is microscopically driven by the creation (publication) of a new CD, as it is the only way in which new edges or new composers can be added into the network (see Figure 3(A)). Since the introduction of CD in the early 1980s, 62,537 CDs have been released, featuring compositions by 13,099 composers, with 131,706 edges connecting CDs and composers. Here we distinguish two different classes of composers, the ones first introduced before 1990 (via the older LPs, many of which were re-released as CDs after 1990) whom we call the oldboys (OBs), and the others first introduced post-1990



whom we call the newbies (NBs). By this criterion, there are 882 oldboys and 13,099 newbies, so that the majority of composers were introduced after the advent of the CD technology. The popularity of the composers, nevertheless, are clearly skewed in the opposite direction: The total degree of the oldboys has increased from 2,553 in 1990 to 89,720 (an increase of 87,167) while only 44,539 links have been created for the newbies, for mean degrees of  $\bar{q}_{OB} = 101.72$  and  $\bar{q}_{NB} = 3.40$ . This implies that, while the number of recorded composers steadily increases by introduction of new ones, the old established ones are perhaps out of their reach in terms of recordings.

#### 4.2 Evolution of bipartite degree distribution indicates predictability for top-ranked composers and explains the rich-get-richer phenomenon in classical music industry

Extending the skewed bipartite degree distribution in the CD-composer network shown in Figure 1(B), we now focus on a more detailed figure of the temporal evolution of  $P(q)$ . In

Figure 3(B) we show the degree distribution of five snapshots of the network taken every five years. The degree distribution approximates a truncated power law that approaches a true power law (with power exponent  $\gamma = 1.89 \pm 0.01$ ) as the network grows. For comparison with another example in culture, in the movie actor network the power exponent is  $\gamma_{\text{actor}} = 2.3$ . The highest-degree (most recorded) composers are WA Mozart, JS Bach, L van Beethoven, F Schubert, and J Brahms throughout the observational period (with final degrees 4,851, 4,292, 3,778, 2,328, and 2,429, respectively). The mean degree of the rest of the composers stays nearly constant at 3.2, again confirming the significant discrepancy between the ‘major minority’ and the ‘minor majority’ in the network landscape of classical music.

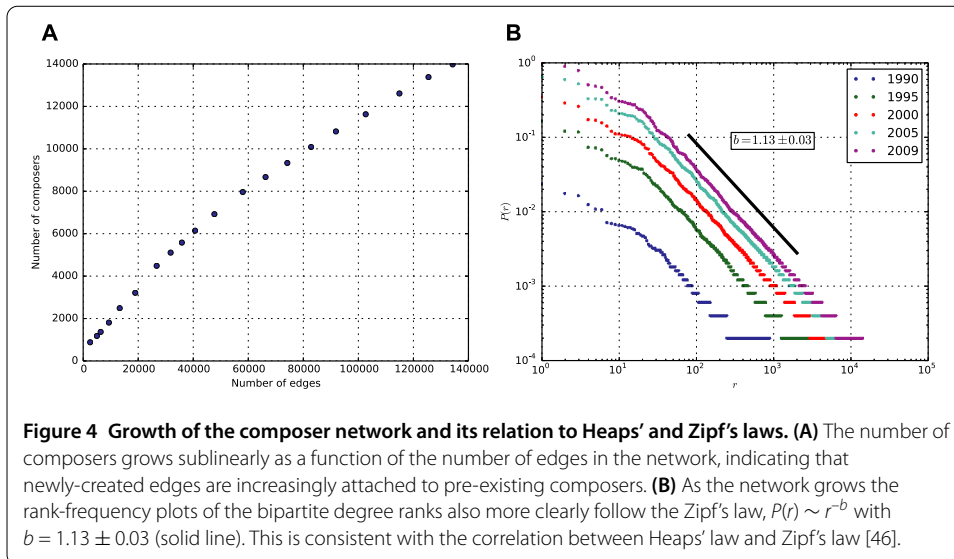
In Figure 3(C) we show the growths of the number of edges in the network and the degrees of six highest-degree composers. The curves appear to be quadratic (i.e.  $\propto t^2$ , dotted curves), suggesting a constant acceleration. Although it remains to be seen if the trend continues, if it does then in the year 2019 WA Mozart would have a bipartite degree of  $q \simeq 10,264$ , JS Bach would have  $q \simeq 9,211$ , and so forth (see Figure 3(C) for all six top composers). Regarding the number of edges, we find that they are becoming increasingly concentrated between the top-degree nodes: The top 1% of the nodes in 1990 (8 composers out of 882) account for 20.3% of all the degrees, while in 2009 the top 1% (139 composers of 13,981) account for 57.1%. These observations tell us that a reasonable growth dynamics of the network must incorporate at minimum two properties, namely, power-law degree distribution and the increasing concentration of degrees on the top-degree nodes.

A popular model of a growing network that exhibits a skewed degree distribution (such as the power law) is the Cumulative Advantage (CA) or Preferential Attachment (PA) [41, 42]. Our CD-composer network, too, boasts typical features that render PA a reasonable mechanism of its growth: It has a skewed distribution and a fixed fraction of the high-degree nodes take up a larger portion of the total degrees as it grows, evidenced by Figures 3(B) and 3(C).

The PA model assumes that the rate at which a node acquires new degrees is proportional to its degree, i.e.  $dk_i/dt = \Pi(k_i)$  with  $\Pi(k_i) \propto k_i$ . This results in a power-law degree distribution with degree exponent 3 [42]. Different  $\Pi(k)$  results in different degree distributions [43, 44]. Often studied is a polynomial form of  $\Pi(k_i) \propto k_i^\alpha$  where  $\alpha > 0$ .  $\alpha = 1$  is the classical PA model. In general,  $\alpha < 1$  results in a stretched exponential degree distribution, and  $\alpha > 1$  results in a more skewed degree distribution than the PA model, a single node linking to all other nodes asymptotically [43]. Given the power exponent of our network,  $\gamma \simeq 2$ , and the increasing share of the degrees by the top nodes, we estimate that for our CD-composer network, effectively  $1 < \alpha < 2$ , which we show as follows. Given the node degrees  $\{q\}$ , we calculate the mean degree  $E_\alpha[q]$  of the node that will take the next added edge under  $\Pi(k) \propto k^\alpha$ , and also the empirical  $E_{\text{data}}[q]$  from data. In Figure 3(D) we plot the quantity

$$\rho \equiv \frac{E_{\text{data}}[q] - E_{\alpha=1}[q]}{E_{\alpha=2}[q] - E_{\alpha=1}[q]}, \tag{1}$$

which shows that in the majority of years our data sits in the range  $0 < \rho < 0.1$ , i.e.  $E_{\alpha=1}[q] < E_{\text{data}}[q] < E_{\alpha=2}[q]$ . Our network indeed shows a superlinear preferential attachment behavior, which can explain the degree distribution and the growing concentration of degree portions on the highest-degree nodes.



**Table 2 Top 10 pairs of composers for edge weights**

Rank	Weight	Pair					
		Name	Period	Active year	Name	Period	Active year
1	680	Giacomo Puccini	Romantic	1891	Giuseppe Verdi	Romantic	1857
2	613	Beethoven	Romantic	1798	WA Mozart	Classical	1773
3	417	Giuseppe Verdi	Romantic	1857	Gaetano Donizetti	Romantic	1822
4	389	Beethoven	Romantic	1798	Johannes Brahms	Romantic	1865
5	384	JS Bach	Baroque	1717	WA Mozart	Classical	1773
6	381	Gioachino Rossini	Romantic	1830	Giuseppe Verdi	Romantic	1857
7	368	JS Bach	Baroque	1717	GF Handel	Baroque	1722
8	355	Giuseppe Verdi	Romantic	1857	WA Mozart	Classical	1773
9	352	Maurice Ravel	Modern	1906	Claude Debussy	Modern	1890
10	350	Franz Schubert	Romantic	1812	WA Mozart	Classical	1773

The weight of a pair indicates the number of records where they are co-featured. The pairs with high weights belong to the same period, or have similar active years.

The behavior of the bipartite degree distribution in Figure 3(B) merits further study. First, in Figure 4(A) we show the number of unique composers in the network as a function of the number of edges in the network. The growth is sublinear, meaning that the rate at which a new composer is added to the network is smaller than the rate of the increase of network edges. This is reminiscent of Heaps' law that shows that the vocabulary grows sublinearly with the document size [45]. Lü *et al.* demonstrated the connection between Zipf's law and Heaps' law [46]; we indeed observe Zipf's law for our system as well via the rank-frequency plot of the bipartite degree in Figure 4(B). Consistent with Figure 3(B), as the network grows larger, Zipf's law becomes clearer.

The sublinear growth in Figure 4(A) indicates that new edges are increasingly attached to pre-existing composers, reflecting the preferential attachment behavior that we discussed above. An interesting question to ask is between which composers (in the projected network) the new edges are created. Each added edge between two composers translates into edge weights (strength), which we believe can also shed on the growth dynamics of the network. In Table 2 we list the top ten strongest composer pairs. Here, too, we see that the newly created edges are placed between composers of similar characteristics: For

example, the strongest pair is Giacomo Puccini (1858-1924) and Giuseppe Verdi (1813-1901); they not only belong to the same Romantic period, but are prominent in a common musical style, Italian opera. Their active years are nearly identical as well. Another interesting example is the pair of Giuseppe Verdi and Gaetano Donizetti (1797-1848). Although Donizetti does not appear on any list of top twenty influential composers on his own in terms of network centrality, he shows a drastic rise in prominence thanks to the number of records shared with a composer (Verdi) with an identical musical style (Romantic Italian opera) and similar active years. Likewise, other top pairs in Table 2 either share the same musical period or show similar active years.

Here we identified two main components of the evolution of the classical composer networks: growth and attachment. We observe that the power-law degree distribution becoming clearer as the network grows, which allows us to forecast the future degrees of the most recorded composers. We also find an effective superlinear preferential attachment behavior which can partially explain the observations.

## 5 Conclusion

In this paper, we studied the network of classical music composers constructed from the comprehensive recording data from ArkivMusic. We presented the basic properties of the CD-composer and composer-composer networks, finding that they exhibit characteristics common to many real-world networks, including the small-world property, the existence of a giant component, high clustering, and heavy-tailed degree distributions. We also explored the global association patterns of composers via centrality, assortative mixing, community structure analyses, which suggest an intriguing interplay between the networks of musicians and our musicological understanding of the western musical tradition in which both are undoubtedly continuously influencing each other. We then examined the growth of the CD-composer bipartite network over time. We identified superlinear preferential attachment as a strong candidate for explaining the increasing concentration of edges around top-degree nodes and the power-law degree distributions. The growth of edges and composer degrees exhibits a quadratic growth, allowing us to forecast the future of several prominent composers. If this behavior persists further into the future, it would suggest an interesting future research direction regarding the growth dynamics of the network.

An analytical investigation as ours on new, large-scale data can provide either new lessons or more rigorous answers to questions that are only partially understood on a subject matter (in our case classical music), although the boundary between the two can often be fuzzy: New findings challenge us to find rigorous answers to already known issues, which in turn can bring about novel discoveries about a system. In our paper we have quantified the overlap (correlation) between manually designated periods and the computationally identified communities; found the superlinear preferential attachment behavior and a quadratic growth of network edges; and showed that the growth of the composer pool is sublinear against the network size, leading to a concentration of edge weights onto specific composer pairs. These findings pose interesting possibilities and opportunities for the type of commercial databases such as the one on classical music that we have used here which, we believe, will play an increasingly important role as a source of many more academic findings than those presented here. We are living in an era where technology is greatly facilitating the consumption of culture by the public, evidenced by the increasing adoption of technology by artistic institutions worldwide for public outreach. This

will lead to larger and higher-quality data that can allow us to learn more about culture and art. Classical music as presented here is merely one of many cultural subjects for this kind of investigation, and we believe that it would be interesting to apply our quantitative methodology to other subjects, e.g. visual arts and literature, and explore their nature in novel ways. Going further, it would also be interesting to compare different systems and find universal characteristics of cultural systems as well as those unique to each. We believe that our work highlights the potential of network science coupled with well-curated large-scale data in answering many pertinent questions.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

Doheum Park and Juyong Park analyzed the data and prepared the figures. All authors wrote and reviewed the manuscript.

#### Author details

<sup>1</sup>Graduate School of Culture Technology and BK21 Plus Postgraduate Organization for Content Science, Korea Advanced Institute of Science & Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea. <sup>2</sup>Arts and Technology Program, University of Texas at Dallas, 800 W. Campbell Road, Richardson, Texas 75080-3021, USA.

#### Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF-20100004910 and NRF-2010-330-B00028), IT R&D program of MSIP/KEIT (10045459), and BK21 Plus Postgraduate Organization for Content Science.

Received: 8 November 2014 Accepted: 9 April 2015 Published online: 22 April 2015

#### References

- O'Hagan J, Borowiecki KJ (2010) Birth location, migration, and clustering of important composers: historical patterns. *Hist Methods* 43(2):81-90
- Malina R, Schich M, Meirelles I (eds) (2014) Arts, humanities, and complex networks. Leonardo ebook series. MIT Press, Cambridge
- Park J (2012) Networks of contemporary popular musicians. *Leonardo* 45(1):78-79
- Schich M, Song C, Ahn Y-Y, Mirsky A, Martino M, Barabási A-L, Helbing D (2014) A network framework of cultural history. *Science* 345(6196):558-562
- Aiello LM, Barrat A, Cattuto C, Schifanella R, Ruffo G (2012) Link creation and information spreading over social and communication ties in an interest-based online social network. *EPJ Data Sci* 1(1):1
- Suárez JL, Sancho F, de la Rosa J (2012) Sustaining a global community: art and religion in the network of Baroque Hispanic-American paintings. *Leonardo* 45(3):281
- Gleiser PM, Danon L (2003) Community structure in jazz. *Adv Complex Syst* 6(04):565-573
- Park J, Celma O, Koppenberger M, Cano P, Buldú JM (2007) The social network of contemporary popular musicians. *Int J Bifurc Chaos* 17(07):2281-2288
- Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854-856
- Uzzi B, Spiro J (2005) Collaboration and creativity: the small world problem. *Am J Sociol* 111(2):447-504
- Newman ME (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167-256
- Jeong H, Néda Z, Barabási A-L (2003) Measuring preferential attachment in evolving networks. *Europhys Lett* 61(4):567
- Serrà J, Corral Á, Boguñá M, Haro M, Arcos JL (2012) Measuring the evolution of contemporary western popular music. *Sci Rep* 2:521
- Bettencourt LM, Lobo J, Helbing D, Kühnert C, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci USA* 104(17):7301-7306
- Serrano MÁ, Boguñá M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci USA* 106(16):6483-6488
- Milgram S (1967) The small world problem. *Psychol Today* 2(1):60-67
- Newman ME (2009) *Networks: an introduction*. Oxford University Press, New York
- Newman ME, Park J (2003) Why social networks are different from other types of networks. *Phys Rev E* 68(3):036122
- Newman ME (2004) Analysis of weighted networks. *Phys Rev E* 70(5):056131
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40:35-41
- Hoppin RH (1978) *Medieval music*. Norton, New York
- Reese G (1959) *Music in the Renaissance*. Norton, New York
- Bukofzer MF (1947) *Music in the Baroque era: from Monteverdi to Bach*. Norton, New York
- Barzun J (1961) *Classic, Romantic, and Modern*, vol 255. University of Chicago Press, Chicago
- Grout DJ, Palisca CV, et al (1996) *A history of Western music*, 5th edn. Norton, New York
- Taruskin R (2009) *The Oxford history of Western music: music in the nineteenth century*, vol 3. Oxford University Press, New York

27. Atlas AW (1998) Renaissance music: music in Western Europe, 1400-1600. Norton, New York
28. Rosen C (1997) The classical style: Haydn, Mozart, Beethoven, vol 1. Norton, New York
29. Kravitt EF (1992) Romanticism today. *Music Q* 76(1):93-109
30. Albright D (2004) Modernism and music: an anthology of sources. University of Chicago Press, Chicago
31. Newman ME (2003) Mixing patterns in networks. *Phys Rev E* 67(2):026126
32. Ahn Y-Y, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. *Nature* 466(7307):761-764
33. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328(5980):876-878
34. Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci* 103(23):8577-8582
35. Sales-Pardo M, Guimera R, Moreira AA, Amaral LAN (2007) Extracting the hierarchical organization of complex systems. *Proc Natl Acad Sci* 104(39):15224-15229
36. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):75-174
37. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):10008
38. Struble JW (1995) The history of American classical music: MacDowell through minimalism. Facts on File, New York
39. Danon L, Diaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. *J Stat Mech Theory Exp* 2005(09):09008
40. Sarigol E, Pfitzner R, Scholtes I, Garas A, Schweitzer F (2014) Predicting scientific success based on coauthorship networks. ArXiv e-prints
41. Simon HA (1955) On a class of skew distribution functions. *Biometrika* 42:425-440
42. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509-512
43. Krapivsky PL, Redner S, Leyvraz F (2000) Connectivity of growing random networks. *Phys Rev Lett* 85(21):4629
44. Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. *Nature* 406(6794):378-382
45. Heaps HS (1978) Information retrieval: computational and theoretical aspects. Academic Press, Orlando
46. Lü L, Zhang Z-K, Zhou T (2010) Zipf's law leads to Heaps' law: analyzing their relation in finite-size systems. *PLoS ONE* 5(12):14139

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---