Metodološki zvezki, Vol. 4, No. 2, 2007, 191-204

Analysis of the Customers' Choice Networks: An Application on Amazon Books and CDs Data

Vladimir Batagelj[1], Nataša Kejžar[2], and Simona Korenjak-Černe[3]

Abstract

Customer's choice implies some kind of relations among products. Customers' choices of products induce a network among them. Analyses of such networks can offer interesting information for marketing. In the paper some network analysis approaches are proposed to analyze such data. Two large networks obtained in 2004 from Amazon Internet bookstore and CD-store are used to illustrate these approaches. All analyses were done with program Pajek.

1 Amazon networks construction

Amazon.com opened its virtual doors in July 1995 with the mission to transform book buying using the Internet into the fastest and easiest way. The Company's principal corporate offices are located in Seattle, Washington. It is one of the leading online shopping sites. It offers huge selection of products, including books, CDs, videos, DVDs, toys and games, electronics, kitchenware, computers etc.

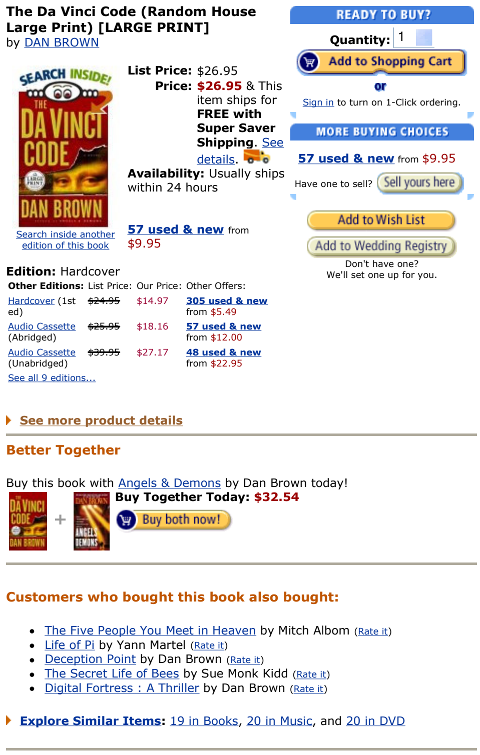
In the Amazon network $\mathcal{N} = (\mathcal{V}, \mathcal{A})$ the **vertices** $\mathcal{V}$ are books/CDs; while the **arcs** $\mathcal{A}$ are determined for each product on the basis of the list of products (books/CDs) in its description under the title: *Customers who bought this book/CD also bought*. The vertex representing a described product is linked with an arc to every product listed in the list. Figure 1 presents an example of the construction of the neighborhood of the book *The Da Vinci Code*.

Using relatively simple program written in Python we 'harvested' the books network from June 16 till June 27, 2004; and the CDs network from July 7 till July 23, 2004. We harvested only the portion of each network reachable from the selected starting book: *Introducing Social Networks* by Michel Forse and Alain Degenne (0761956042) / starting CD: *The Pros and Cons of Hitchhiking* by Roger Waters (B0000025ZF).

[1] University of Ljubljana, Faculty of Mathematics and Physics, Department of Mathematics; Vladimir.Batagelj@fmf.uni-lj.si

[2] University of Ljubljana, Faculty of Social Sciences, Department of Informatics and Methodology; Natasa.Kejzar@fdv.uni-lj.si

[3] University of Ljubljana, Faculty of Economics, Department of Statistics; Simona.Cerne@ef.uni-lj.si



The Da Vinci Code (Random House Large Print) [LARGE PRINT]
by DAN BROWN

Search inside another edition of this book

List Price: $26.95
Price: **$26.95** & This item ships for **FREE with Super Saver Shipping**. See details.
Availability: Usually ships within 24 hours

57 used & new from $9.95

Quantity: 1

or

Sign in to turn on 1-Click ordering.

57 used & new from $9.95

Have one to sell?

Don't have one?
We'll set one up for you.

Edition: Hardcover

Other Editions:	List Price:	Our Price:	Other Offers:
Hardcover (1st ed)	~~$24.95~~	$14.97	**305 used & new** from $5.49
Audio Cassette (Abridged)	~~$25.95~~	$18.16	**57 used & new** from $12.00
Audio Cassette (Unabridged)	~~$39.95~~	$27.17	**48 used & new** from $22.95

See all 9 editions...

See more product details

Better Together

Buy this book with Angels & Demons by Dan Brown today!
Buy Together Today: $32.54

Customers who bought this book also bought:

- The Five People You Meet in Heaven by Mitch Albom (Rate it)
- Life of Pi by Yann Martel (Rate it)
- Deception Point by Dan Brown (Rate it)
- The Secret Life of Bees by Sue Monk Kidd (Rate it)
- Digital Fortress : A Thriller by Dan Brown (Rate it)

Explore Similar Items: 19 in Books, 20 in Music, and 20 in DVD

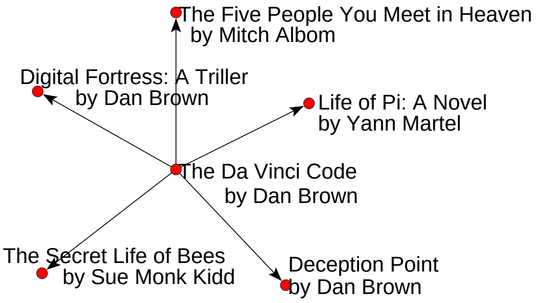


Figure 1: Dan Brown: *The Da Vinci Code.*

2 Description of obtained networks

The books network has 216737 vertices (books) and 982296 arcs. The CDs network has 79244 vertices (CDs) and 526271 arcs. By construction both networks have limited out-degree and are weakly connected. 178281 books have the maximum out-degree 5; and 55373 CDs have the maximum out-degree 8. Figure 2 shows the distributions of books/CDs by their in-degrees. The book with largest in-degree 553 is *Dan Brown: The Da Vinci Code*. The CDs with largest in-degree are *The Shins: Chutes Too Narrow* (706), and *Norah Jones: Feels Like Home* (675).

2.1 Strong components

The books network has 1787 nontrivial strong components, the largest of size 198808. The CDs network has 237 nontrivial strong components. The number of strong components strongly decreases with their size as can be noticed in Table 1. There are 130 strong components of size 2, 18 strong components of size 3 etc. But there is only 1 component of size 39, 1 component of size 84, 1 component of size 207, and 1 component of size 73928, and these are the only components with the sizes larger than 30.

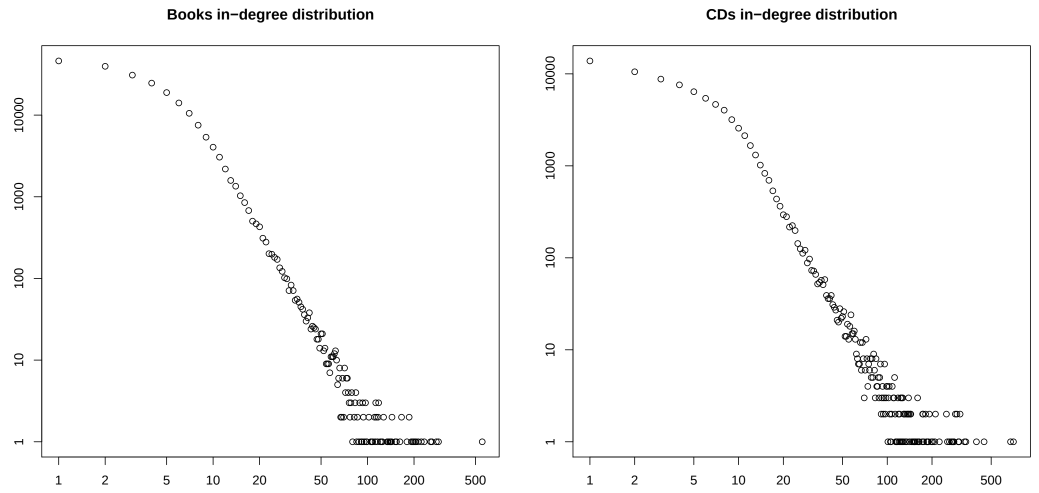


Figure 2: In-degree distributions.

Table 1: Distribution of strong components by their size in CDs network.

```
[1]       3512  130  18  7  2  1  1  3  5  5
[11]         4   12   5  5  9  6  0  3  2  3
[21]         1    3   2  2  2  1  1  0  0  0
[39]         1
[84]         1
[207]        1
[73928]      1
```

2.2 Symmetrical subnetworks

To obtain undirected networks from the collected directed books/CDs networks two approaches were used:

- **network skeleton**: transform each arc into an edge and delete multiple edges;
- **symmetrical subnetwork**: replace pairs of opposite arcs with edges and delete the remaining arcs. Vertices with degree 0 are removed.

In the symmetrical subnetwork two vertices are linked by an edge if and only if there exist arcs in both directions. These vertices are worthy to be considered in detail to find the topic-purchase flow. Therefore we decided to analyze symmetrical subnetwork as an undirected network.

The symmetrical subnetwork on books has 186113 vertices, 218563 edges and 18967 components – the largest of size 59289. The second largest component is much smaller with 294 books and the third one includes 257 books.

The symmetrical subnetwork on CDs has 69708 vertices, 124834 edges and 3012 components – the largest of size 51692. The second largest component includes only 102 CDs and the third one 97.

3 Analysis

The goal of our analysis was to **identify the important parts** of both Amazon networks and to uncover their characteristics. To determine the most popular books and CDs the **hubs and authorities procedure** (Kleinberg, 1998) was used, and to **determine the main topics** in the networks the **islands approach** (Batagelj and Zaveršnik, 2004) was applied.

3.1 Hubs and authorities

A vertex is a **good hub** if it points to many good authorities, and it is a **good authority** if it is pointed by many good hubs.

To formalize this idea each vertex v in the network gets two weights x_v and y_v. The corresponding vectors x and y are related by equations $x = W^T y$ and $y = Wx$, where $W = [w_{uv}]$ is the weight matrix of the network. It can be proved that x and y are the principal eigenvectors of matrices $W^T W$ and WW^T (Kleinberg, 1998).

In the Amazon books/CD network out-degree is truncated, therefore hubs cannot reach values as large as authorities. Figure 3 shows hubs (white) and authorities (gray) around the main authority *The Da Vinci Code* in the books network. Black vertices are both good hubs and good authorities. Figure 4 shows five hubs and five authorities with the largest weights around the main authority *The Shins: Chutes Too Narrow* in the CDs network. The sizes of the vertices correspond to their weights.

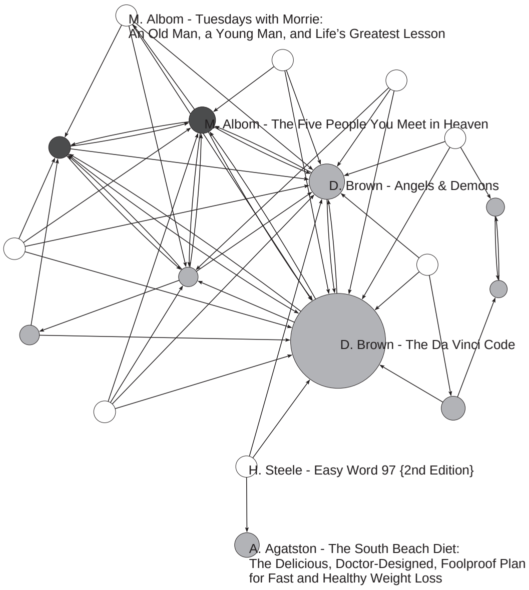

M. Albom - Tuesdays with Morrie:
An Old Man, a Young Man, and Life's Greatest Lesson
M. Albom - The Five People You Meet in Heaven
D. Brown - Angels & Demons
D. Brown - The Da Vinci Code
H. Steele - Easy Word 97 {2nd Edition}
A. Agatston - The South Beach Diet:
The Delicious, Doctor-Designed, Foolproof Plan
for Fast and Healthy Weight Loss

Figure 3: Hubs and authorities around *The Da Vinci Code* in books network.

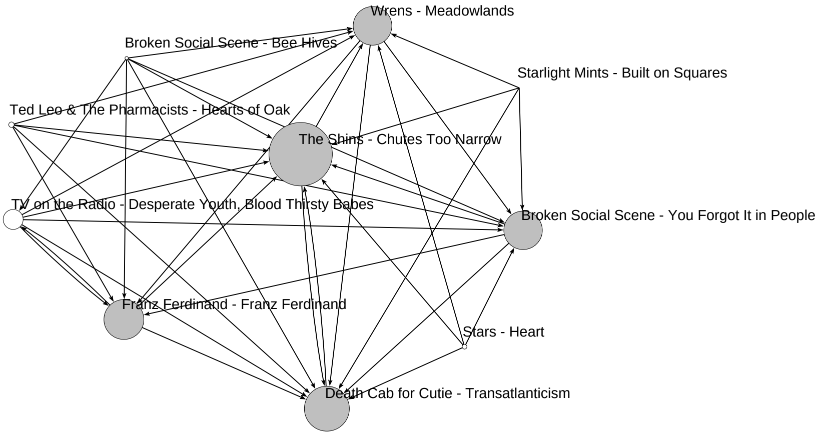

Wrens - Meadowlands
Broken Social Scene - Bee Hives
Starlight Mints - Built on Squares
Ted Leo & The Pharmacists - Hearts of Oak
The Shins - Chutes Too Narrow
TV on the Radio - Desperate Youth, Blood Thirsty Babes
Broken Social Scene - You Forgot It in People
Franz Ferdinand - Franz Ferdinand
Stars - Heart
Death Cab for Cutie - Transatlanticism

Figure 4: Hubs and authorities around *The Shins: Chutes Too Narrow* in CDs network.

3.2 Islands

Islands are connected parts of a network containing locally the most important vertices/ lines with respect to a given property/weight. Formally:

For a given network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$ with a vertex property p a **regular vertex island** is a cluster $\mathcal{C}$ of vertices for which the corresponding induced subgraph is connected, and the property values of vertices in the cluster $\mathcal{C}$ are larger than the values of vertices in the cluster's neighborhood $N(\mathcal{C})$:

$$\max_{u \in N(\mathcal{C})} p(u) < \min_{v \in \mathcal{C}} p(v)$$

The set of vertices is a **local vertex peak**, if it is a regular vertex island and all of its vertices have the same value. Vertex island with a single local vertex peak is called a **simple vertex island**.

For a given network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$ with a line weight w a **regular line island** is determined by a cluster $\mathcal{C}$ of vertices in which exists a spanning tree $\mathcal{T}$ on $\mathcal{C}$ such that the weights of lines with exactly one endpoint in the cluster are smaller than the weights of lines of the tree:

$$\max_{e(u:v) \in \mathcal{L}:(u \in \mathcal{C} \wedge v \notin \mathcal{C})} w(e) < \min_{e \in \mathcal{L}(\mathcal{T})} w(e)$$

where $e(u : v)$ denotes a line e linking vertex u with vertex v.

Similarly as simple vertex island we define simple line island as an island with only one peak.

We analyzed Amazon networks using several vertex properties and line weights. Here we present some results obtained with **clustering coefficient** (Watts and Strogatz, 1998) as a vertex property on CDs network and **triangular count** as a line weight on books network, although all presented approaches can be used on both networks.

3.2.1 Vertex property – clustering coefficient

The clustering coefficient measures the local density of a network in given vertex. It is defined as a proportion of links between the vertices within the neighborhood of the vertex by the number of all possible links in its neighborhood. Let $\deg(v)$ denote the degree of vertex v, $|\mathcal{L}(G_1(v))|$ the number of lines among vertices ($\mathcal{L} = \mathcal{A}$ for arcs and $\mathcal{L} = \mathcal{E}$ for edges) in 1-neighborhood of vertex v, and Δ the maximum vertex degree in a network.

Clustering coefficient $CC_1(v)$ of vertex v is defined as follows:

$$\begin{array}{lcl} \text{for a \textit{directed} network:} & CC_1(v) & = & \dfrac{|\mathcal{A}(G_1(v))|}{\deg(v)\cdot(\deg(v)-1)} \\ \text{for an \textit{undirected} network:} & CC_1(v) & = & \dfrac{2|\mathcal{E}(G_1(v))|}{\deg(v)\cdot(\deg(v)-1)} \end{array}$$

The problem with the clustering coefficient CC_1 is that it has high values on vertices of small degree. To neutralize this effect we propose to use in data analytic tasks the **corrected clustering coefficient**:

$$CC_1'(v) = \frac{\deg(v)}{\Delta} CC_1(v)$$

Directed CDs network

In the directed network on CDs are 4415 simple vertex islands for $p(v) = CC_1'(v)$. The distribution of vertex islands by their size (number of vertices) is presented in the Table 2.

Table 2: Distribution of vertex islands by their size in the directed CDs network.

[1]	1625	576	356	273	221	154	176	152	158	144
[11]	128	105	87	72	38	34	21	19	12	19
[21]	13	8	5	2	5	1	2	2	0	0
[31]	2	1	2	1	0	0	1	0	0	0

The island of the maximal size contains 37 CDs. The analysis of CDs within each island shows that they are either from the same author or they belong to the same type of music. For example the seven largest vertex islands based on corrected clustering coefficient in the directed network on CDs, can be identified as:

#of CDs	Description
37	Barbra Streisand
34	Modern Scandinavian folk music
33	Soul and blues
33	Hed Kandi's house and disco music
32	Progressive rock music
31	Hip-hop
31	Julio Iglesias

Undirected CDs network

In the undirected network (symmetrical subnetwork) on CDs are 8302 simple vertex islands based on the corrected clustering coefficient $p(v) = CC_1'(v)$. The distribution of vertex islands by their size is presented in the Table 3.

Table 3: Distribution of vertex islands by the size in the undirected CDs network.

```
[1]   1229  2158  1237  769  520  441  380  360  332  296
[11]   184   141    76   61   35   24   12   15    9    4
[21]     5     7     3    1    0    1    1    0    0    1
```

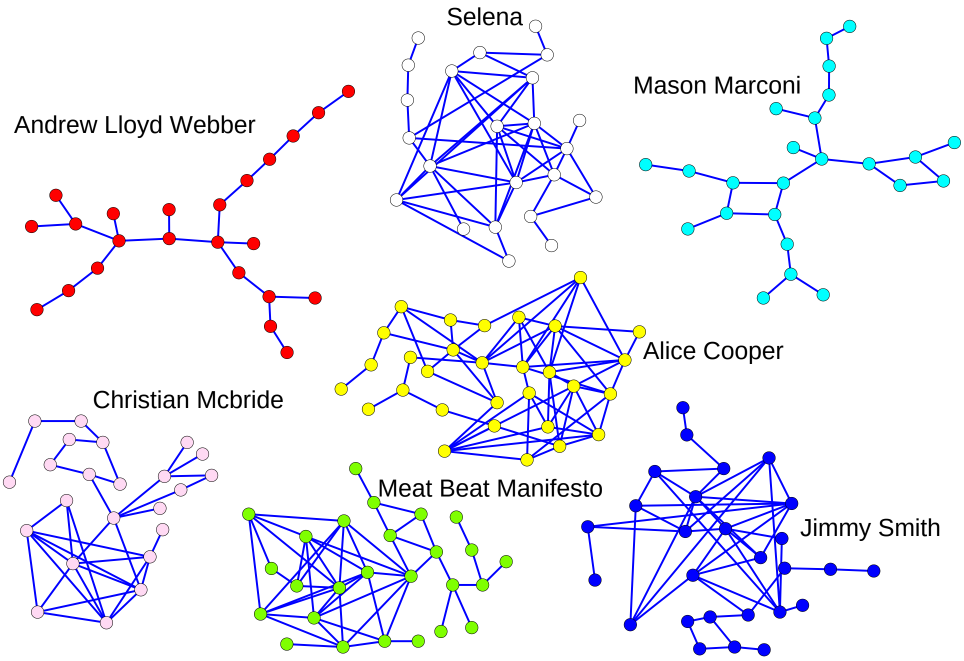


Figure 5: Seven largest vertex islands in the undirected network on CDs.

Also in these islands the CDs inside each island are either from the same author or they belong to the same type of music. The island with the maximal size contains 30 CDs. The seven largest vertex islands based on the corrected clustering coefficient in the undirected network on CDs, shown in the Figure 5, can be identified as:

#of CDs	Author(s) or type of music
30	Alice Cooper
27	Jimmy Smith, Eddie Harries, Lou Rawls
26	Meat Beat Manifesto, L.S.G.
24	Erotic Fantasy (Mason Marconi, unknown artists)
23	Musicals (Andrew Lloyd Webber and others)
23	Jazz (Christian Mcbride, Nicholas Payton, Mark Whitfield)
23	Selena

The vertex island that includes CDs with songs from well known **musicals** such as *Sunset Boulevard* and *Starlight Express* from Andrew Lloyd Webber, *Cabaret* and

Chicago from John Kander, and *My One and Only* from George and Ira Gershwin is presented in detail separately in the Figure 6.

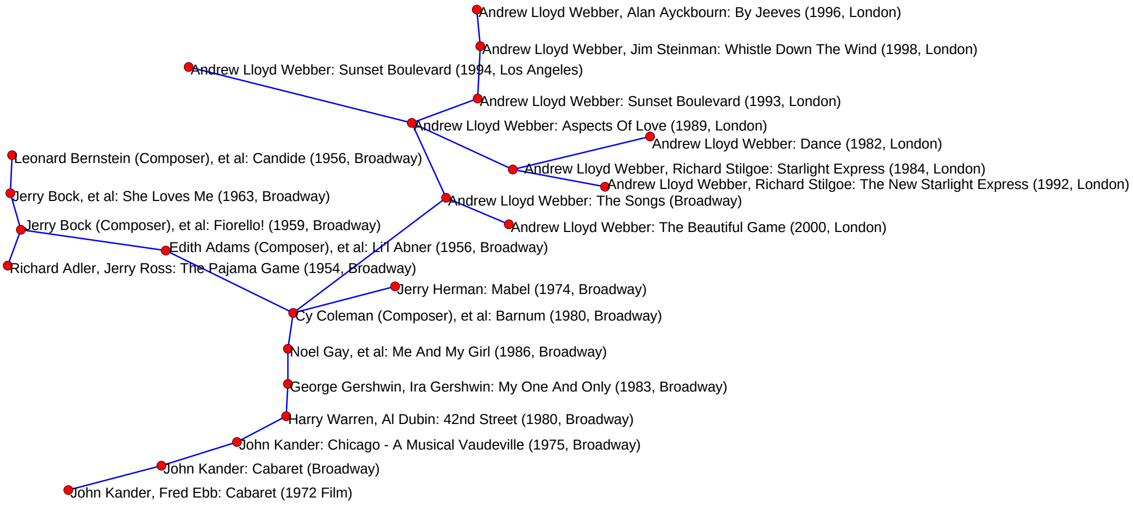

Andrew Lloyd Webber, Alan Ayckbourn: By Jeeves (1996, London)
Andrew Lloyd Webber, Jim Steinman: Whistle Down The Wind (1998, London)
Andrew Lloyd Webber: Sunset Boulevard (1994, Los Angeles)
Andrew Lloyd Webber: Sunset Boulevard (1993, London)
Andrew Lloyd Webber: Aspects Of Love (1989, London)
Andrew Lloyd Webber: Dance (1982, London)
Leonard Bernstein (Composer), et al: Candide (1956, Broadway)
Andrew Lloyd Webber, Richard Stilgoe: Starlight Express (1984, London)
Andrew Lloyd Webber, Richard Stilgoe: The New Starlight Express (1992, London)
Jerry Bock, et al: She Loves Me (1963, Broadway)
Andrew Lloyd Webber: The Songs (Broadway)
Jerry Bock (Composer), et al: Fiorello! (1959, Broadway)
Andrew Lloyd Webber: The Beautiful Game (2000, London)
Edith Adams (Composer), et al: Li'l Abner (1956, Broadway)
Richard Adler, Jerry Ross: The Pajama Game (1954, Broadway)
Jerry Herman: Mabel (1974, Broadway)
Cy Coleman (Composer), et al: Barnum (1980, Broadway)
Noel Gay, et al: Me And My Girl (1986, Broadway)
George Gershwin, Ira Gershwin: My One And Only (1983, Broadway)
Harry Warren, Al Dubin: 42nd Street (1980, Broadway)
John Kander: Chicago - A Musical Vaudeville (1975, Broadway)
John Kander: Cabaret (Broadway)
John Kander, Fred Ebb: Cabaret (1972 Film)

Figure 6: Vertex island: Musicals.

Not only the size of the island, but also the weight of vertices inside the island is worth considering, because it contains more detailed information about the importance of the vertex. In the undirected CDs network 9 vertices have the highest weight (equal to 1) and all of them are in the same vertex island presented in the Figure 7. They form a clique of order 9. The island can be described as 'Fred Astaire and Ella Fitzgerald island'. Six CDs have the second largest weight. All of them are in the same vertex island with 4 other CDs that are related with the Beatles. This island is presented in the Figure 8.

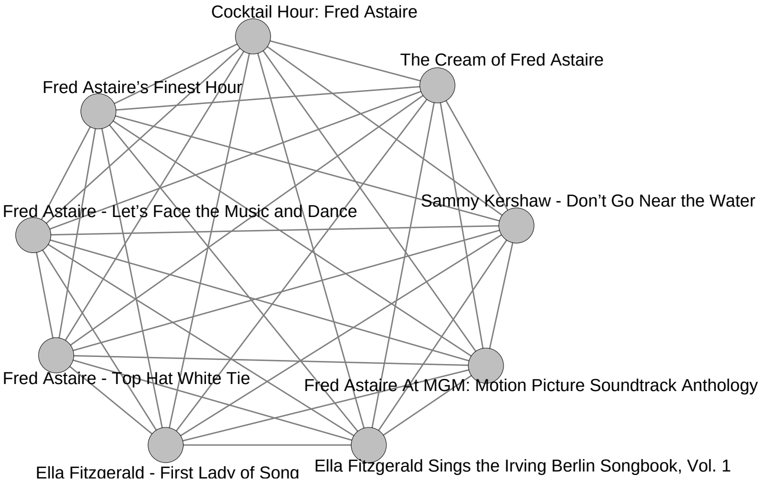

Cocktail Hour: Fred Astaire
The Cream of Fred Astaire
Fred Astaire's Finest Hour
Sammy Kershaw - Don't Go Near the Water
Fred Astaire - Let's Face the Music and Dance
Fred Astaire - Top Hat White Tie
Fred Astaire At MGM: Motion Picture Soundtrack Anthology
Ella Fitzgerald - First Lady of Song
Ella Fitzgerald Sings the Irving Berlin Songbook, Vol. 1

Figure 7: Vertex island with the largest weights.

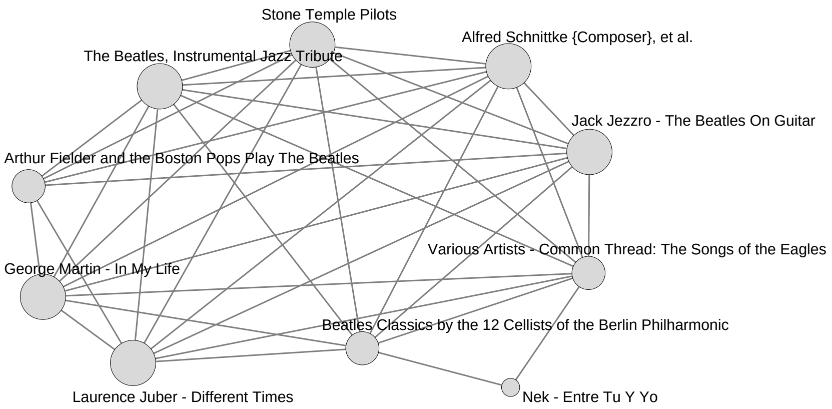

Stone Temple Pilots
Alfred Schnittke {Composer}, et al.
The Beatles, Instrumental Jazz Tribute
Jack Jezzro - The Beatles On Guitar
Arthur Fielder and the Boston Pops Play The Beatles
Various Artists - Common Thread: The Songs of the Eagles
George Martin - In My Life
Beatles Classics by the 12 Cellists of the Berlin Philharmonic
Laurence Juber - Different Times
Nek - Entre Tu Y Yo

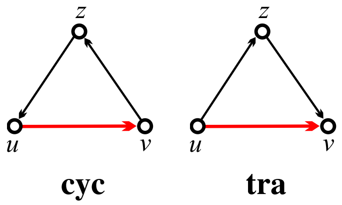
Figure 8: Vertex island with the second largest weights.

3.2.2 Line weight – triangle count

Triangular weight $w_3(e)$ of a line e counts the number of triangles to which the line belongs. If the line e belongs to a k-clique then $w_3(e) \geq k - 2$. Therefore the use of triangle count enables efficient detection of dense parts of networks. This approach can be used on:

- *undirected networks*, to get edge weights
- *directed networks*, to get arc weights

In the directed network two basic types of triangles can be observed: cyclic and transitive.



Books network as directed network

Figure 9 presents the distributions of the size of line islands in the directed networks of books and CDs considering the number of cyclic triangles as arcs' weights.

Examining the books included in the islands, the same or similar topics can be noticed inside the same island. Figure 10 shows line (arc) islands with at least 25 books and the main topics of the books inside them. Two of these islands are presented separately: the island including novels written by the **same author** Catherine Cookson in Figure 11, and the island including books about the **same topic** precious stones in Figure 12.

Books network as undirected network

Figure 13 shows line (edge) islands with at least 35 books in the symmetric subnetwork

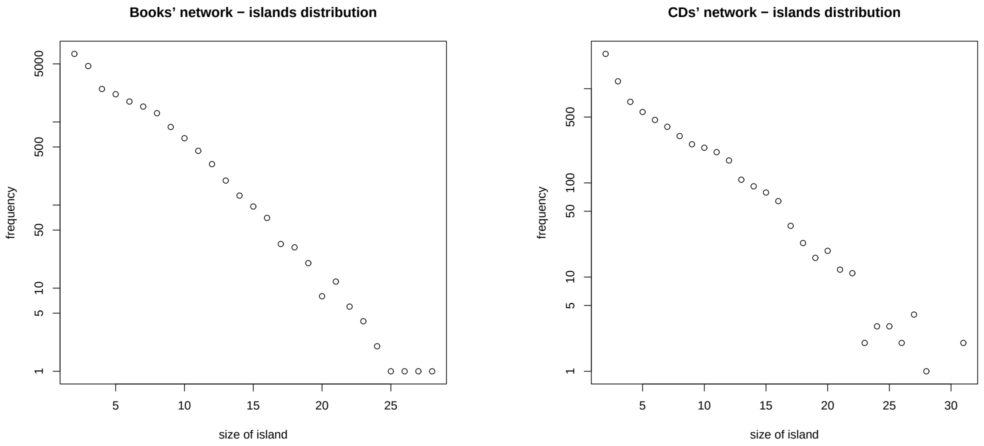


Figure 9: Simple line (arc) islands distribution by their size. Note the log-linear scale of the graphs, which shows the sharp drop in frequency of larger line islands.

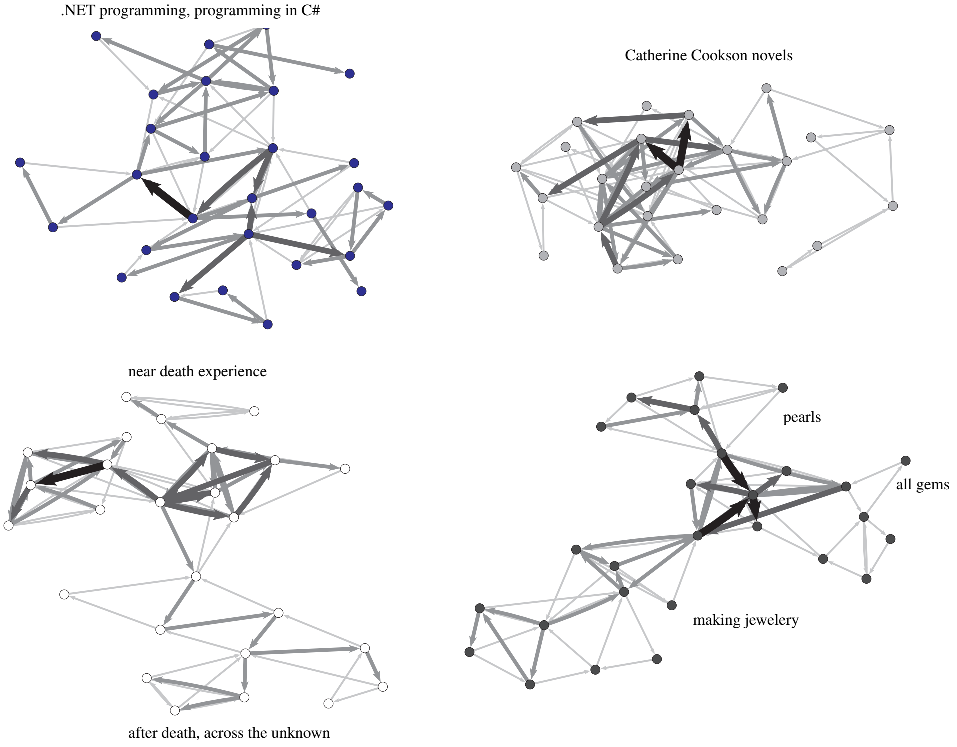


Figure 10: Arc islands with at least 25 vertices.

of the books network. Chaining is observed. We conjecture that the chaining is the 'backbone' of the topic, because two vertices in this network are connected only if there exist arcs both ways between them. These vertices therefore have to be really important for topic-purchase flow. Two of these islands are presented separately: the island including

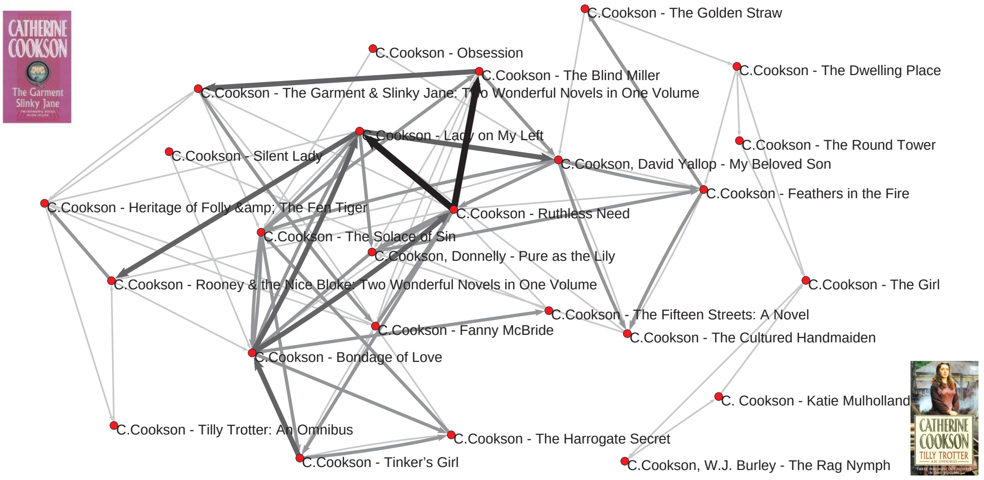

C.Cookson - The Golden Straw
C.Cookson - Obsession
C.Cookson - The Dwelling Place
C.Cookson - The Blind Miller
C.Cookson - The Garment & Slinky Jane: Two Wonderful Novels in One Volume
C.Cookson - Lady on My Left
C.Cookson - The Round Tower
C.Cookson - Silent Lady
C.Cookson, David Yallop - My Beloved Son
C.Cookson - Feathers in the Fire
C.Cookson - Heritage of Folly &amp; The Fen Tiger
C.Cookson - Ruthless Need
C.Cookson - The Solace of Sin
C.Cookson, Donnelly - Pure as the Lily
C.Cookson - Rooney & the Nice Bloke: Two Wonderful Novels in One Volume
C.Cookson - The Girl
C.Cookson - The Fifteen Streets: A Novel
C.Cookson - Fanny McBride
C.Cookson - The Cultured Handmaiden
C.Cookson - Bondage of Love
C. Cookson - Katie Mulholland
C.Cookson - Tilly Trotter: An Omnibus
C.Cookson - The Harrogate Secret
C.Cookson - Tinker's Girl
C.Cookson, W.J. Burley - The Rag Nymph
CATHERINE COOKSON
The Garment
Slinky Jane
CATHERINE COOKSON
TILLY TROTTER

Figure 11: Island of Catherine Cookson novels.

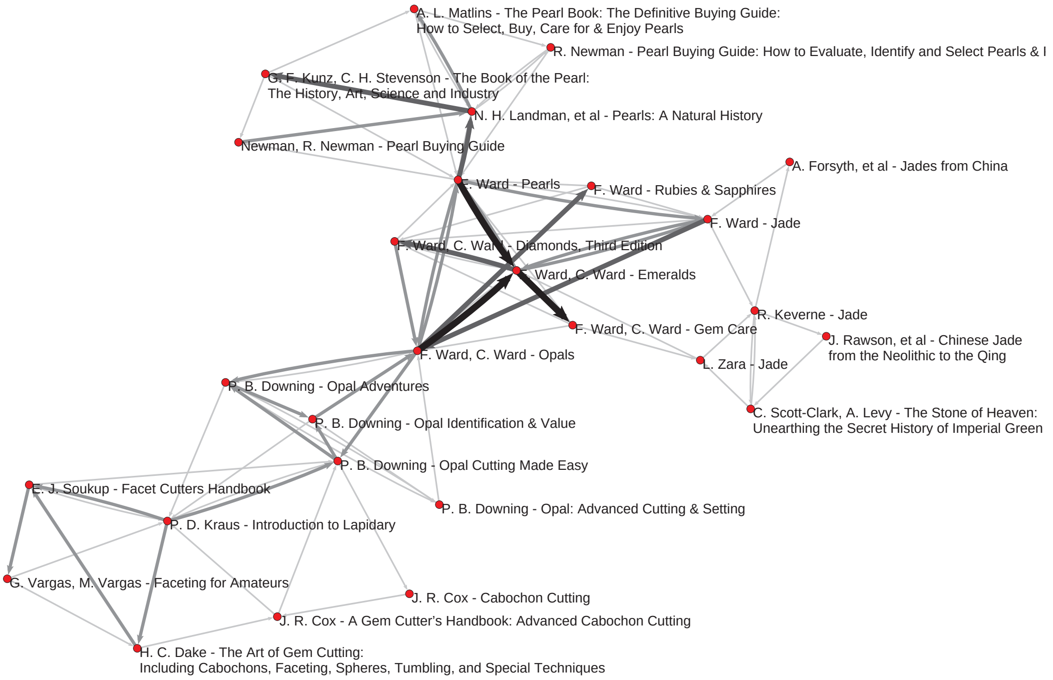

A. L. Matlins - The Pearl Book: The Definitive Buying Guide:
How to Select, Buy, Care for & Enjoy Pearls
R. Newman - Pearl Buying Guide: How to Evaluate, Identify and Select Pearls & I
G. F. Kunz, C. H. Stevenson - The Book of the Pearl:
The History, Art, Science and Industry
N. H. Landman, et al - Pearls: A Natural History
Newman, R. Newman - Pearl Buying Guide
A. Forsyth, et al - Jades from China
F. Ward - Pearls
F. Ward - Rubies & Sapphires
F. Ward - Jade
F. Ward, C. Ward - Diamonds, Third Edition
F. Ward, C. Ward - Emeralds
R. Keverne - Jade
F. Ward, C. Ward - Gem Care
J. Rawson, et al - Chinese Jade
from the Neolithic to the Qing
F. Ward, C. Ward - Opals
L. Zara - Jade
R. B. Downing - Opal Adventures
C. Scott-Clark, A. Levy - The Stone of Heaven:
Unearthing the Secret History of Imperial Green
P. B. Downing - Opal Identification & Value
P. B. Downing - Opal Cutting Made Easy
E. J. Soukup - Facet Cutters Handbook
P. B. Downing - Opal: Advanced Cutting & Setting
P. D. Kraus - Introduction to Lapidary
G. Vargas, M. Vargas - Faceting for Amateurs
J. R. Cox - Cabochon Cutting
J. R. Cox - A Gem Cutter's Handbook: Advanced Cabochon Cutting
H. C. Dake - The Art of Gem Cutting:
Including Cabochons, Faceting, Spheres, Tumbling, and Special Techniques

Figure 12: Island of precious stones.

books for children in Figure 14, and the island including books for students of literature in Figure 15.

Since the number of vertices in the island is not the only characteristic that has to be considered in the analysis, we inspected also the weights on lines. They range from 0 to 4, which implies that this network is relatively sparsely connected (there are at most 4 triangles above each line, and they represent only 0.5 % of all the edges). There are 73 islands (of 2 or more vertices) with at least one line weight 4, they are mostly of size 2 or 3 (68 of them) and 5 islands with 6 vertices, which are complete graphs. Four of them consist of books of the same author(s): (1) Will Durant and Ariel Durant (history books), (2) Peter J. D'Adamo and Catherine Whitney (books about food and blood type), (3) Immanuel Velikovsky (history and legend books) and (4) Rachel Rubin Wolf (a book series). The last one of the islands consists of books of the Silva mind control method, where only 2 books are written by Silva himself.

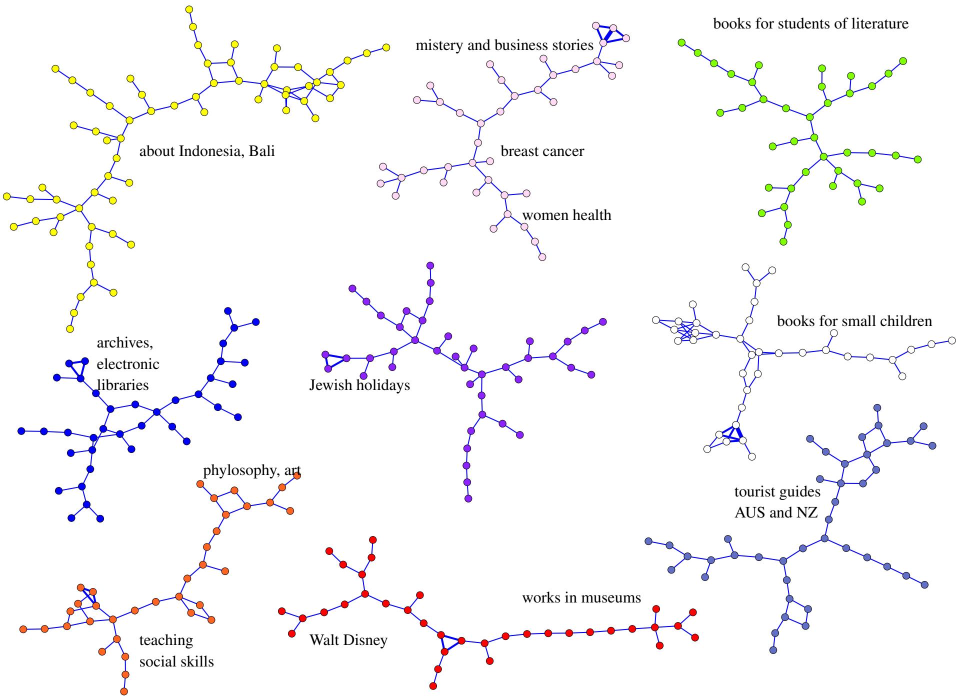


Figure 13: Edge islands with more than 35 vertices.

4 Discussion

Like well known data mining techniques (Berry and Linoff, 2004) used to extract information from large amounts of data, approaches from network analysis can show many interesting connections among products. In the paper some of these approaches have been shown for customers' choice networks constructed from the Amazon Internet bookstore, based on the list of products (books/CDs) under the title: *Customers who bought this book/CD also bought*.

With the hubs and authorities procedure (Kleinberg, 1998) the most important products were identified. By determining different types of islands (Batagelj and Zaveršnik, 2004) the important topics of books/CDs were detected. For the weights of vertices (books/CDs) the corrected clustering coefficient as a measure of relative local density in a given vertex was chosen. For line weights the number of triangles was chosen which also enables efficient detection of dense parts of networks. Due to large amount of results we presented only some of them to show the main possibilities of such analysis. To obtain the insight into the logic of groups formation, their characteristics have to be analyzed.

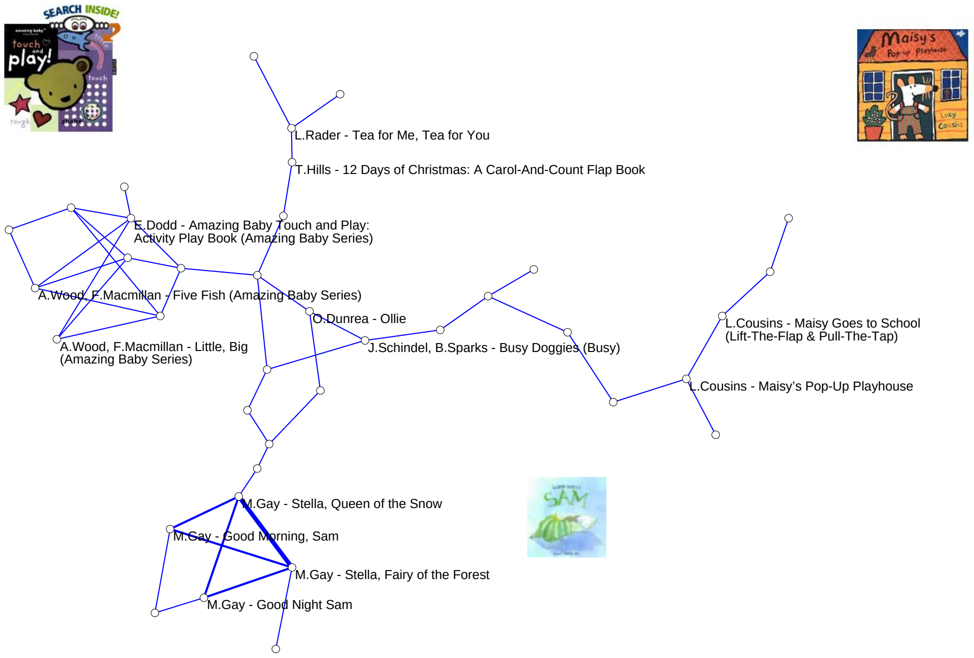

L.Rader - Tea for Me, Tea for You
T.Hills - 12 Days of Christmas: A Carol-And-Count Flap Book
E.Dodd - Amazing Baby Touch and Play:
Activity Play Book (Amazing Baby Series)
A.Wood, F.Macmillan - Five Fish (Amazing Baby Series)
O.Dunrea - Ollie
A.Wood, F.Macmillan - Little, Big
(Amazing Baby Series)
J.Schindel, B.Sparks - Busy Doggies (Busy)
L.Cousins - Maisy Goes to School
(Lift-The-Flap & Pull-The-Tap)
L.Cousins - Maisy's Pop-Up Playhouse
M.Gay - Stella, Queen of the Snow
M.Gay - Good Morning, Sam
M.Gay - Stella, Fairy of the Forest
M.Gay - Good Night Sam

Figure 14: Island of books for small children.

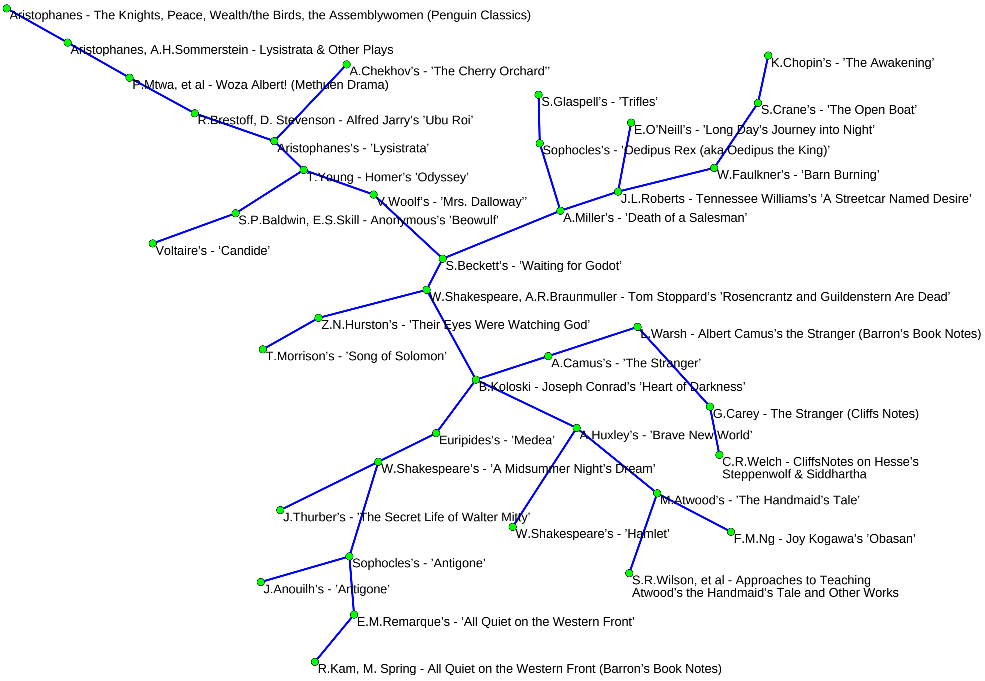

Aristophanes - The Knights, Peace, Wealth/the Birds, the Assemblywomen (Penguin Classics)
Aristophanes, A.H.Sommerstein - Lysistrata & Other Plays
P.Mtwa, et al - Woza Albert! (Methuen Drama)
A.Chekhov's - 'The Cherry Orchard''
R.Brestoff, D. Stevenson - Alfred Jarry's 'Ubu Roi'
Aristophanes's - 'Lysistrata'
T.Young - Homer's 'Odyssey'
V.Woolf's - 'Mrs. Dalloway''
S.P.Baldwin, E.S.Skill - Anonymous's 'Beowulf'
Voltaire's - 'Candide'
S.Glaspell's - 'Trifles'
K.Chopin's - 'The Awakening'
S.Crane's - 'The Open Boat'
E.O'Neill's - 'Long Day's Journey into Night'
Sophocles's - 'Oedipus Rex (aka Oedipus the King)'
W.Faulkner's - 'Barn Burning'
J.L.Roberts - Tennessee Williams's 'A Streetcar Named Desire'
A.Miller's - 'Death of a Salesman'
S.Beckett's - 'Waiting for Godot'
W.Shakespeare, A.R.Braunmuller - Tom Stoppard's 'Rosencrantz and Guildenstern Are Dead'
Z.N.Hurston's - 'Their Eyes Were Watching God'
L.Warsh - Albert Camus's the Stranger (Barron's Book Notes)
T.Morrison's - 'Song of Solomon'
A.Camus's - 'The Stranger'
B.Koloski - Joseph Conrad's 'Heart of Darkness'
G.Carey - The Stranger (Cliffs Notes)
Euripides's - 'Medea'
A.Huxley's - 'Brave New World'
W.Shakespeare's - 'A Midsummer Night's Dream'
C.R.Welch - CliffsNotes on Hesse's
Steppenwolf & Siddhartha
M.Atwood's - 'The Handmaid's Tale'
J.Thurber's - 'The Secret Life of Walter Mitty'
W.Shakespeare's - 'Hamlet'
F.M.Ng - Joy Kogawa's 'Obasan'
Sophocles's - 'Antigone'
J.Anouilh's - 'Antigone'
S.R.Wilson, et al - Approaches to Teaching
Atwood's the Handmaid's Tale and Other Works
E.M.Remarque's - 'All Quiet on the Western Front'
R.Kam, M. Spring - All Quiet on the Western Front (Barron's Book Notes)

Figure 15: Island of books for students of literature.

Some of them can be obtained from article descriptions from Amazon; the others can be deduced from additional sources.

We believe that analysis of customers' choice networks based on customer's purchases can offer interesting information for marketing, for example in advertising (interesting themes, most popular books, most popular authors etc.).

Another interesting theme would also be to observe how networks are changing through time.

Acknowledgment

This work was supported by the Ministry of Higher Education, Science and Technology of Slovenia, Project J1-6062. It is a detailed version of the talks presented at the conference *Methodology and Statistics*, September 16–18, 2004, Ljubljana, Slovenia.

References

[1] Batagelj, V. (2004): Collecting network data from the Amazon in Python. `http://vlado.fmf.uni-lj.si/pub/networks/data/econ/amazon/amazon.htm`

[2] Batagelj, V. and Zaveršnik, M. (2004): Islands – identifying themes in large networks. Presented at Sunbelt XXIV Conference, Portorož, May 2004.

[3] Berry, M. J. A. and Linoff, G. S. (2004): *Data Mining Techniques*. Second Edition. Wiley Publishing, Inc.

[4] Kleinberg, J. (1998): Authoritative sources in a hyperlinked environment, *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*.

[5] De Nooy, W., Mrvar, A., and Batagelj, V. (2005): *Exploratory Social Network Analysis with Pajek*. New York: Cambridge University Press.

[6] Zaveršnik, M. (2003): *Razčlembe omrežij (Network decompositions)*. PhD. Thesis, FMF, University of Ljubljana.

[7] Watts, D. J. and Strogatz, S. (1998): Collective dynamics of 'small-world' networks. *Nature*, **393**. 440-442.

[8] The Amazon Internet Store:
`http://www.amazon.com/`

[9] The Pajek program – home page:
`http://vlado.fmf.uni-lj.si/pub/networks/pajek/`

[10] The Python Programming Language:
`http://www.python.org/`