

Full Length Research Paper

A demonstration of correlation graphs to human body dimensions

Carel Nicolaas Bezuidenhout* and Ryan Rhett Domleo

School of Engineering, Private Bag X01, Scottsville, 3209, South Africa.

Accepted 16 July, 2013

A large number of variables and multivariate analyses are required to study complexity. Network analysis is also a valuable complex systems analysis tool. Recently techniques were developed where correlation matrixes are presented as networks, known as correlation graphs. These have been used mainly to study structures in time series data, often in the financial stock markets, and correlation graphs have not been widely adopted. The aim of this paper is to demonstrate correlation graphs as a general research tool by using intuitively understandable data. Correlation graphs of human body shape dimensions were generated and noise reducing techniques are demonstrated. The results support the underlying structures of human biology and gender is distinguished with an accuracy of 89%. Weight plays a vital role and height is more influential in males than in females. With age, males and females become more alike. The potential application of the correlation graphing methodology appears extensive.

Key words: Correlation graph, asset graph, multivariate analyses, graph theory, body shape

INTRODUCTION

Complexity in nature, society and science is often associated with a large number of relatively independent processes or entities. At the same time a certain degree of feedback and correlation between entities exist and the system as a whole exhibits behaviour at an aggregated level, which cannot be explained by merely studying the individual parts (Newman, 2011). It is therefore not surprising that a large number of variables over the entire system are measured when complex systems are studied (Simeonov et al., 2003; Singh et al., 2004). Consequently, there is a need to (a) make sense of these large datasets, (b) detect and quantify aggregated system structures, and (c) understand the smaller elements inside the system.

Multivariate analysis is a well established statistical discipline and many different techniques exist to detect, display and analyse predictability relationships between variables. These include covariate analyses (Slot et al., 2010), multivariate regression (Clark and Harper, 2008), canonical and principal component analysis (Bulluck et al., 2002; Zhang et al., 2005), discriminant analysis (Zhang et al., 2006; Ferris et al., 2008), hierarchical cluster analysis, such as dendrograms (Gil et al., 2008; Ye and Wright, 2010; Yuan et al., 2008) and self-organising maps (Ballabio et al., 2013).

One of the most well-established empirical methods to study linear relationships between different variables is the century old Pearson Correlation Coefficient (r), and

hence the Pearson's Correlation Matrix when more than two variables are simultaneously considered. Ke et al. (2012) note rightfully that "(Pearson's Correlation Coefficient) is shown to be one of the most desirable correlation measures for its ability to capture the departure of two variables from independence". All the tools mentioned above remain to some extent limited when depicting a large dataset that may contain hundreds of variables and that exhibits complex behaviour at an aggregated level.

Graph Theory, or network analyses, is a well-established discipline and has become a valuable tool to make sense of large interconnected systems, such as the internet (Cuomo et al., 2012), ecosystems (Larsen et al., 2012) and time series data (Mantegna, 1999). Often a data matrix can be depicted in graph form and a multitude of potentially valuable descriptors, such as betweenness and planarity (Newman, 2011) exist from which aggregated information pertaining to the graph topology could be extracted. In recent years statistical physicists have developed valuable techniques whereby a correlation matrix is presented as a network and where vertexes (nodes) depict individual variables and edges (lines) depict the correlation strength between variables (Heimo, 2008; Onnela, 2004, 2003). Costa et al. (2011) identify correlation mapping as one of six categories in the network analysis arena, but in the literature to date, correlation graphs have been used almost exclusively to study the structure in time series data and often with a strong focus on the financial stock markets (e.g. Yoshikawa et al., 2012). Surprisingly, most correlation graph literature to date has been primarily published in statistical-physics journals with little adoption of this potentially powerful methodology into other scientific disciplines. Bezuidenhout et al. (2012) used correlation graphs to study soil health and compared the methodology to Principle Component Analysis.

The aim of this paper is to demonstrate the potential value of correlation graphs in science by analysing a familiar data set, while showing how this methodology can be used to unpack some of the complexities of a real world system. The correlation graph methodology is relatively simple compared to other multivariate techniques and is supported by a strong visualisation component. The data that were used in this study are familiar and readers will be able to intuitively relate to the findings. This paper makes a number of contributions; (a) unlike most other related studies, this research focuses on non-time series related data, (b) for the first time the dataset was normalised and transposed to produce a new correlation graph projection, and (c) we also introduce two novice techniques to address redundancy, namely maximum two (MAX2) and minimum delta (min δ).

METHODOLOGY

In a paper that focuses on statistics education, Heinz et al. (2003) present a dataset of 24 body shape attributes for 247 men and 260 women. The dataset (T), which is a 24×507 matrix is described in Table 1 and has been made available on the internet (Heinz et al., 2003). Measurements are explained in more detail by Heinz et al. (2003). Each of the variables ($T_{1, 2, \dots, 24}$) was normalised to possess a mean of zero and a standard deviation of one. Normalisation does not affect the correlation coefficient (r) between variables. The data were used to calculate a 24×24 Pearson correlation matrix between variables. Because variables were normalised, the data matrix can be transposed (T^T) and a 507×507 Pearson correlation matrix between individual persons ($T^T_{1, 2, \dots, 507}$) can also be calculated.

Gómez et al. (2009) promote the use of the unsigned value of the correlation coefficient $|r|$ to determine the degree of predictability between two variables. This approach applies well to the 24×24 body shape variable correlation matrix where even a strong inverse correlation still implies a strong relationship between two variables. However, in the 507×507 correlation matrix between individual persons, a strong inverse correlation between two persons would imply that these subjects are quite different from each other, compared to a strong positive correlation. The unsigned 24×24 and signed 507×507 correlation matrixes were subsequently represented by correlation graphs with 24 and 507 vertexes, respectively. All vertexes were connected by edges where the weight of an edge between two particular vertexes depicts the strength of the respective correlation value between the two variables. These networks have 276 and 128,271 weighted edges, respectively.

Heimo et al. (2008) and Souma et al. (2009) note that the correlation matrix could be used to determine the underlying network structure of a complex system, but due to the relatively large number of edges, it also contains a certain degree of redundancy that needs to be removed. Gómez et al. (2009) discuss techniques to remove the edges that represent relatively weak correlations between variable pairs without losing valuable structural information. Several techniques exist to reduce the number of edges in a correlation graph. Gómez et al. (2009) removed all edges with weights below a certain threshold value (e.g. $r < 0.3$). In this case the proportion (p) of correlation coefficients that are included in the graph is not specified, and the specific threshold value that was chosen could be perceived as a subjective decision. In contrast, the Pareto principle may suggest that only the highest 20% ($p=0.2$) of correlation coefficients should be included in the graph (Bezuidenhout et al., 2012). While this technique may be appropriate in systems where there is an exponential decline in correlation strengths, it is also likely to fail in many applications. Figure 1, for example, illustrates a case where the strength of correlation coefficients does not decline logarithmically. A maximum spanning tree (MST) approach is more sophisticated (Mantegna, 1999) and will remove all the redundancy without disconnecting any vertexes from the network, while simultaneously retaining the strongest overall correlation structure.

We now present two more objective approaches to potentially remove redundant edges from the correlation graph. First, we include only the two highest correlation coefficients per variable. This was named the MAX2 correlation graph. Secondly, we sort the correlation coefficients from the highest to the lowest and plot them against p , as presented in Figure 1. In searching for the fewest number of correlation coefficients that would describe the largest number of important relationships in the system, we calculate the Euclidean distance (δ) between each data point and the origin of the graph. The closest data point to the origin (as depicted by an

Table 1. Body shape variables that were used in this study (Heinz et al., 2003).

Abbreviation	Description
<i>Biacr</i>	Biacromial diameter (width between shoulder joints)
<i>Biili</i>	Biiliac diameter (breadth of pelvis)
<i>Bitro</i>	Bitrochanteric diameter (width between top of legs below hip joints)
<i>ChestDe</i>	Chest depth between spine and sternum at nipple level*
<i>ChestDi</i>	Chest diameter at nipple level*
<i>Elbow</i>	Elbow diameter, sum of two elbows
<i>WristDi</i>	Wrist diameter, sum of two wrists
<i>KneeDi</i>	Knee diameter, sum of two knees
<i>AnkleDi</i>	Ankle diameter, sum of two ankles
<i>Shoul</i>	Shoulder girth over deltoid muscles
<i>ChestGi</i>	Chest girth*(males at the nipple line; females just above breast tissue)
<i>Waist</i>	Waist girth, narrowest part below rib cage, average between contracted and relaxed
<i>Navel</i>	Navel (or "Abdominal") girth at umbilicus and iliac crest, iliac crest as a landmark
<i>Hip g</i>	Hip girth at level of bitrochanteric diameter
<i>Thigh</i>	Thigh girth below gluteal fold ^Δ
<i>Bicep</i>	Bicep girth, flexed ^Δ , average between right and left girths
<i>Forea</i>	Forearm girth, extended, palm up ^Δ
<i>KneeGi</i>	Knee girth over patella, slightly flexed position ^Δ
<i>Calf</i>	Calf maximum girth ^Δ
<i>AnkleGi</i>	Ankle minimum girth ^Δ
<i>WristGi</i>	Wrist minimum girth ^Δ
<i>Age</i>	Age
<i>Weigh</i>	Weight
<i>Heigh</i>	Height

*measurement performed at mid-expiration; ^Δaverage between right and left.

arrow in Figure 1) represents the minimum δ threshold ($\text{min-}\delta$) and all the correlation coefficients to the left of this point are included in the correlation graph. It can be argued that diminishing returns apply if correlation coefficients to the right of $\text{min-}\delta$ were to be included in the correlation graph. In this case variables that have an overall low correlation with all other variables may be completely disconnected from the network, likewise, well correlated variables will have a high degree of connectivity. The $\text{min-}\delta$ value presents an opportunity to potentially quantify and compare the nature of different correlating systems. A relatively low $\text{min-}\delta$ value will, for example, imply that only a few correlation coefficients are sufficient to describe a system, which may suggest that the system may be inherently simple compared to a system with a higher $\text{min-}\delta$ value.

The correlation graphs that were generated above can now be plotted and subjected to the well-established Kamada-Kawai energising algorithm (Kamada and Kawai, 1989). This algorithm determines an equilibrium position for each vertex based on the weights and connectivity of its edges. It produces a visual map of the system that can be interpreted with relative ease. The Pajek networking software (De Nooy et al., 2012) was used to perform these tasks. A wide range of additional topological analytical techniques are available (Newman, 2011) to further help unpack the nature of the system. There is a significant research opportunity to relate different topological features of a correlation graph to the nature and complexity of the real system represented.

RESULTS

Table 2 reflects the Pearson correlation matrix between 24 different body measurements. The correlation coefficients are graphically represented in Figure 1 and it is evident that the overall correlation structure between variables is relatively strong. There are no strong inverse correlations present in the data. This supports a sizing phenomenon where, overall, larger people have larger measurements in almost all attributes. Seventeen of the 24 variables correlated strongly with at least one other variable in the dataset ($r \geq 0.80$). The Age variable has the lowest degree of correlation. All these trends were also confirmed by Principal Component Analyses.

Figure 2 illustrates a maximum spanning tree (MST) of the 507 samples based on the strongest overall correlation structure. Yellow and blue vertexes depict males and females, respectively. Darker edges depict higher correlation values. The size of each vertex reflects the age of the particular person. Displayed images (a)-(q) of different body shapes were taken from Baek and Lee

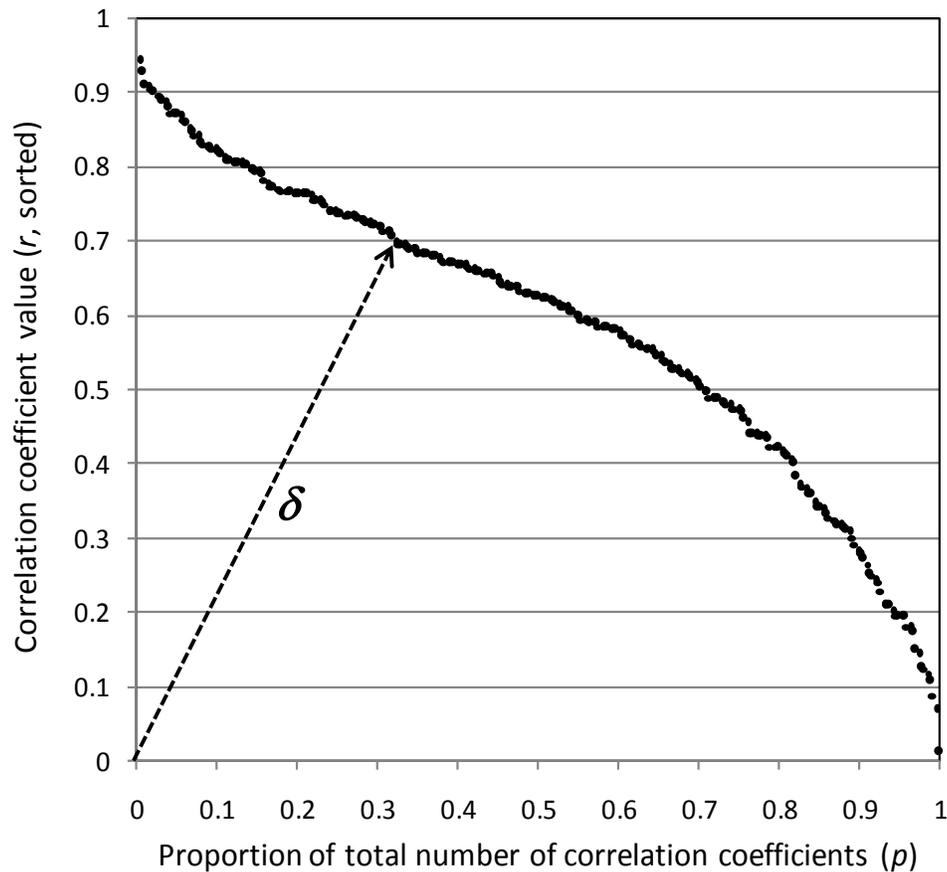


Figure 1. Correlation coefficients (r) from a 24×24 human body shape correlation matrix sorted and plotted against their rank number and expressed as a fraction (p).

(2012), courtesy of Elsevier Publishers, and were matched with specific persons in the dataset as indicated by alphabetical symbols.

The MST reveals a strong underlying structure of similarities between groups of people. Although Age was identified as a relatively weak predictor in the correlation matrix (Table 2), it appears to drive a significant degree of clustering within the MST. Ninety-two of the 506 edges in Figure 2 interconnect people who fall within the oldest age quartile. The binomial probability for this occurrence is lower than 0.0000001%. Certain human shape categories seem to be almost exclusively reserved for older people. Older females cluster around F1, while older males cluster around two shape categories (M1 and M2). The shape images (e) and (f) display two male profiles that exclusively belong to older persons in this dataset. Likewise, images (m) and (n) are female profiles that are more associated with younger individuals.

There are several methods available to partition a graph based on clustering algorithms (De Nooy et al.,

2012) in order to identify more uniform sub-groups within the dataset. A red line in Figure 2 indicates where the MST was split into two groups. This was done by separating the network at the vertex with the highest betweenness value. The two sub-groups, named M and F, contain predominantly males and females, respectively. Statistics of these two groups are provided in Table 3. The MST differentiates between males and females with an accuracy of 88.95%.

Figure 3(a) and (b) illustrate the $\min\text{-}\delta$ correlation graphs for body measurements for the M and F groups, respectively. Vertex colours depict different body areas. Statistics for these graphs are presented in Table 4. In general a large number of body measurements correlate strongly, which is confirmed by the density of edges in Figures 3(a) and (b). This makes it more difficult to draw conclusions from the $\min\text{-}\delta$ correlation graphs. Although it is not depicted in this paper, the equivalent $\min\text{-}\delta$ graph of the 507×507 correlation matrix produced more clusters and sub-structures. The $\min\text{-}\delta$ value of 0.70 in

Table 2. Pearson correlation matrix between human body dimension variables.

Variables	Biacr	Biili	Bitro	ChestDe	ChestDi	Elbow	WristDi	KneeDi	AnkleDi	Shoul	ChestGi	Waist	Navel	Hip g	Thigh	Bicep	Forea	KneeGi	Calf	AnkleGi	WristGi	Age	Weigh	Heigh
Biacr	1.00																							
Biili	0.31	1.00																						
Bitro	0.49	0.67	1.00																					
ChestDe	0.58	0.36	0.47	1.00																				
ChestDi	0.77	0.33	0.52	0.67	1.00																			
Elbow	0.77	0.32	0.53	0.67	0.76	1.00																		
WristDi	0.72	0.28	0.47	0.61	0.73	0.84	1.00																	
KneeDi	0.64	0.44	0.61	0.55	0.66	0.73	0.71	1.00																
AnkleDi	0.66	0.37	0.50	0.60	0.67	0.82	0.77	0.72	1.00															
Shoul	0.79	0.28	0.48	0.74	0.87	0.82	0.78	0.68	0.69	1.00														
ChestGi	0.72	0.33	0.49	0.81	0.87	0.80	0.77	0.65	0.71	0.93	1.00													
Waist	0.64	0.43	0.57	0.80	0.79	0.69	0.68	0.62	0.64	0.82	0.88	1.00												
Navel	0.31	0.58	0.62	0.62	0.50	0.44	0.40	0.47	0.44	0.52	0.62	0.75	1.00											
Hip g	0.34	0.56	0.75	0.56	0.52	0.44	0.42	0.58	0.41	0.53	0.58	0.69	0.83	1.00										
Thigh	0.12	0.41	0.53	0.36	0.31	0.21	0.19	0.43	0.19	0.32	0.36	0.42	0.60	0.83	1.00									
Bicep	0.70	0.30	0.48	0.73	0.79	0.80	0.76	0.68	0.69	0.90	0.91	0.80	0.56	0.56	0.41	1.00								
Forea	0.75	0.29	0.48	0.72	0.81	0.86	0.81	0.72	0.74	0.89	0.89	0.78	0.49	0.51	0.35	0.94	1.00							
KneeGi	0.51	0.47	0.62	0.56	0.59	0.59	0.58	0.73	0.54	0.62	0.61	0.66	0.61	0.73	0.64	0.62	0.66	1.00						
Calf	0.51	0.41	0.59	0.55	0.60	0.58	0.58	0.69	0.54	0.63	0.61	0.63	0.52	0.67	0.63	0.64	0.67	0.80	1.00					
AnkleGi	0.60	0.34	0.54	0.59	0.64	0.66	0.65	0.65	0.68	0.68	0.67	0.66	0.52	0.58	0.42	0.67	0.71	0.74	0.76	1.00				
WristGi	0.77	0.26	0.48	0.68	0.76	0.85	0.86	0.73	0.76	0.84	0.82	0.73	0.44	0.46	0.24	0.85	0.90	0.64	0.65	0.75	1.00			
Age	0.09	0.25	0.27	0.32	0.19	0.20	0.21	0.17	0.24	0.18	0.25	0.37	0.42	0.23	-0.02	0.18	0.15	0.12	0.11	0.14	0.19	1.00		
Weigh	0.73	0.50	0.67	0.80	0.83	0.80	0.76	0.77	0.73	0.88	0.90	0.90	0.71	0.76	0.56	0.87	0.87	0.80	0.77	0.76	0.82	0.21	1.00	
Heigh	0.75	0.38	0.49	0.55	0.63	0.74	0.68	0.59	0.69	0.67	0.62	0.55	0.31	0.34	0.12	0.59	0.66	0.53	0.45	0.57	0.69	0.07	0.72	1.00

the F Group compared to 0.59 in the M group suggests that body shapes of persons in the F group are more proportional and predictable. In contrast, the M Group contains more noise and depicts a wider variety of relationships between

body measurements. In most cases variables that describe the same body area (e.g. yellow torso vertexes) cluster together. It is, however, interesting to note that the diameter of limb joints, such as the elbow, wrist and knee, often correlate

better with each other compared to other limb measurements in the same body region. In both cases the centrality and degree of the *Weigh* variable proposes that a large number of body measurements are strongly regulated by weight

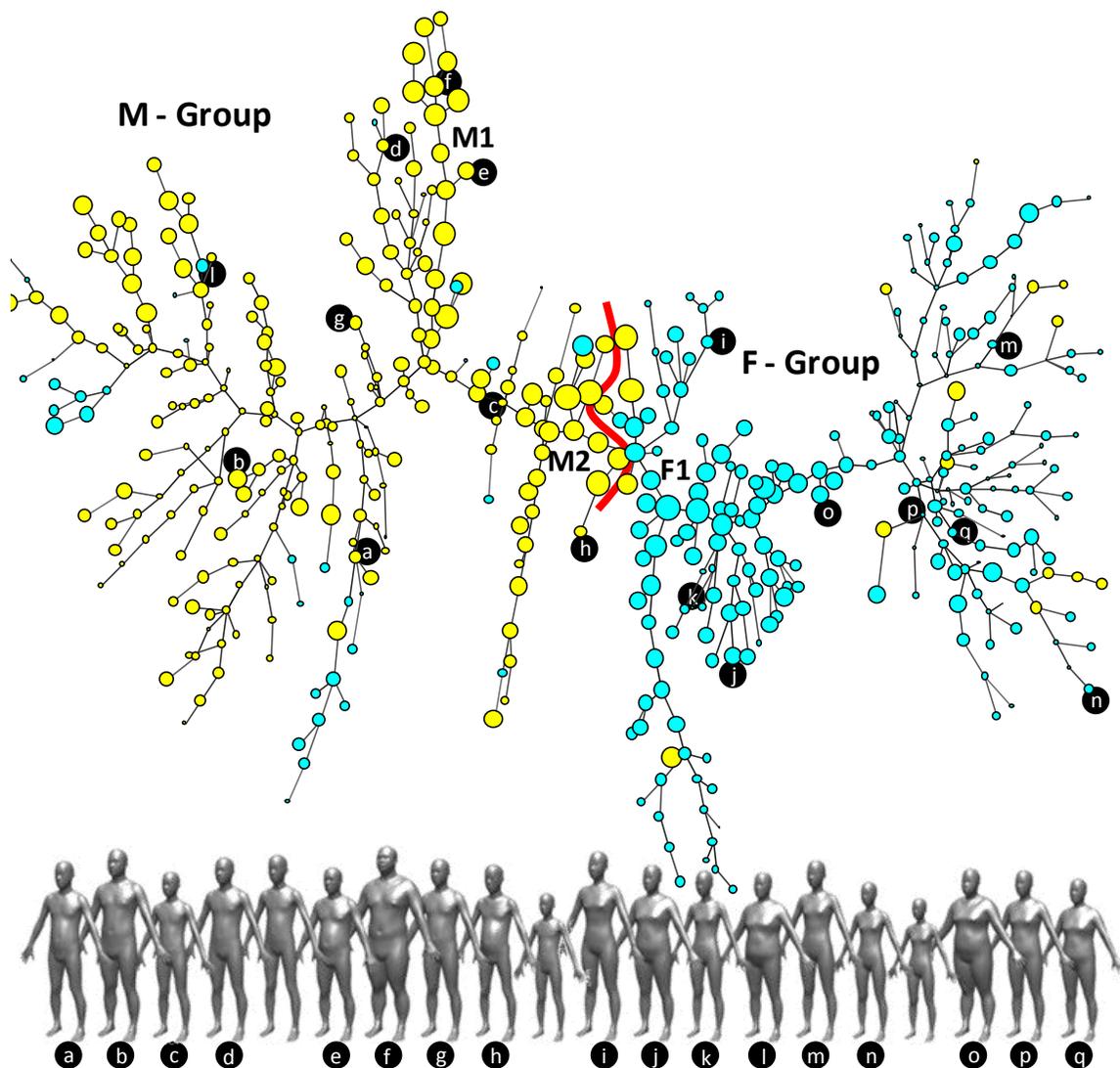


Figure 2. A maximum spanning tree of correlation coefficients between 507 people based on 24 body shape measurements. Images of body shapes were used from Baek and Lee (29) with permission from Elsevier Publishers.

Table 3. Basic statistics of the M Group and F Group derived from the MST of 507 people.

Group	No. of males	No. of females	Total
M Group	226	35	261
F Group	21	225	246
Total	247	260	507

breaking out into a separate network. The M Group, in particular is interesting. Upper limb measurements correlate with the upper torso, while the lower torso measurements correlate with age. Intuitively this makes

sense because males who develop their upper limbs are likely to also develop the features of their upper torso. The knee and ankle diameters remain interesting since they continue to correlate better with joints in the upper limbs, rather than with other lower limb features. This suggests that the diameter of joints in the limbs, in both the M and F Groups, are probably more related to genetics than to lifestyle and age.

DISCUSSION

Body shapes of males and females can normally be

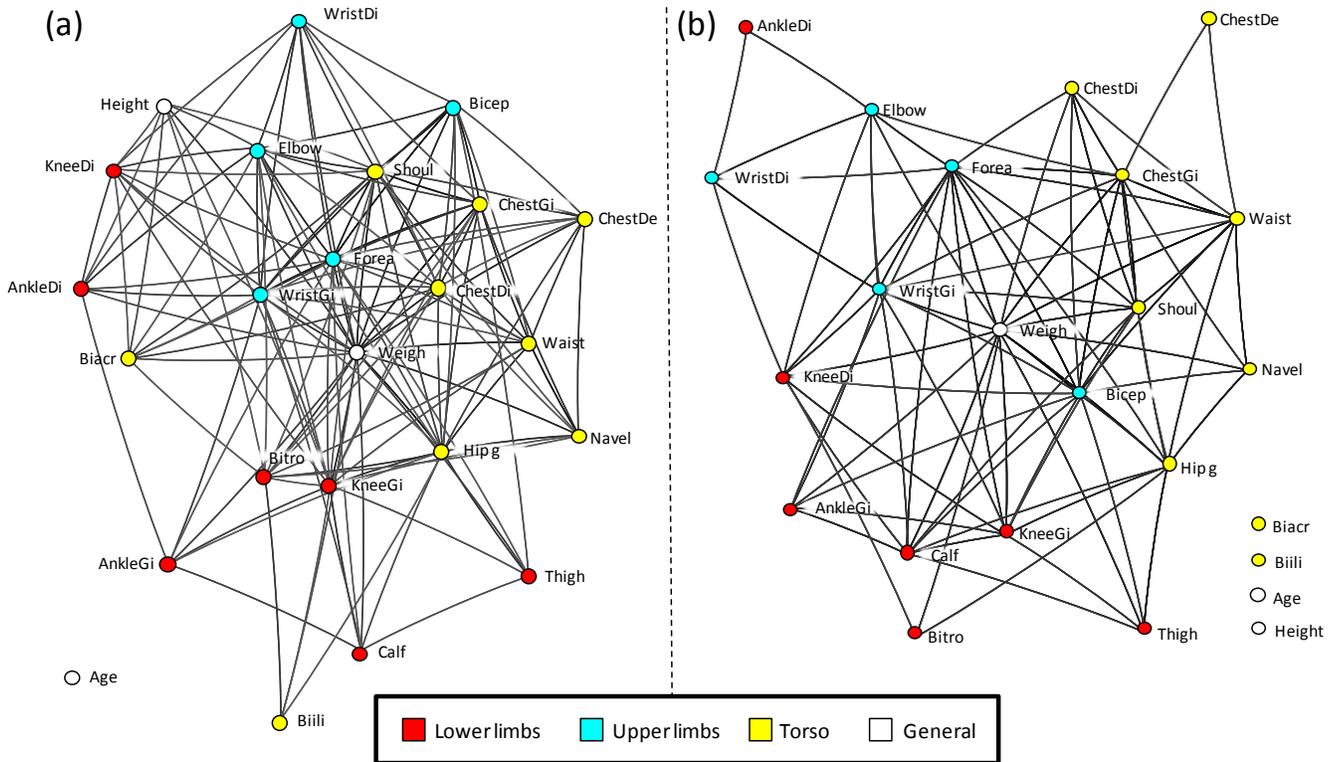


Figure 3. Correlation graphs of the (a) M Group and (b) F Group after redundant edges were removed by using the $\text{min-}\delta$ approach.

Table 4. Statistical attributes of data and correlation graphs presented in this paper.

Correlation Graph	<i>n</i>	Edges	$\text{min-}\delta$	Vertex with the highest degrees
MST : 507 persons	507	506	0.35	
$\text{min-}\delta$: M Group	24	138	0.59	<i>Weigh</i> (22 edges)
$\text{min-}\delta$: F Group	24	84	0.70	<i>Weigh</i> (16 edges)
MAX2: M Group	24	37		<i>Weigh</i> (9 edges)
MAX2: F Group	24	41		<i>Weigh</i> (11 edges)
MAX2: M Group (<i>Weigh</i> excluded)	23	29		<i>Elbow, Shoul</i> (5 edges)
MAX2: F Group (<i>Weigh</i> excluded)	23	28		<i>Elbow, ChestGi, Navel</i> (4 edges)

distinguished with ease. Weight plays a vital role in shaping body measurements and height is more influential in males compared to females. Older males and females often become more alike and this is often associated with changes in the lower torso. Males have an increased ability to change their upper torso and upper limbs disproportionately to the rest of their bodies. Generally the attributes that describe the shape of the torso and those that describe the limbs are better correlated among each other compared to across body

regions. The diameter of joints in a person’s limbs (e.g. elbows, knees and ankles) are genetically determined and are less dependent on lifestyle, weight and age. Intuitively, the results that were generated in this paper appear correct. However, in an unfamiliar dataset, these patterns could easily go undetected when a conventional correlation matrix, multiple regression analysis or Principal Component Analysis is applied.

This paper demonstrates how a correlation graph in conjunction with the maximum spanning tree algorithm

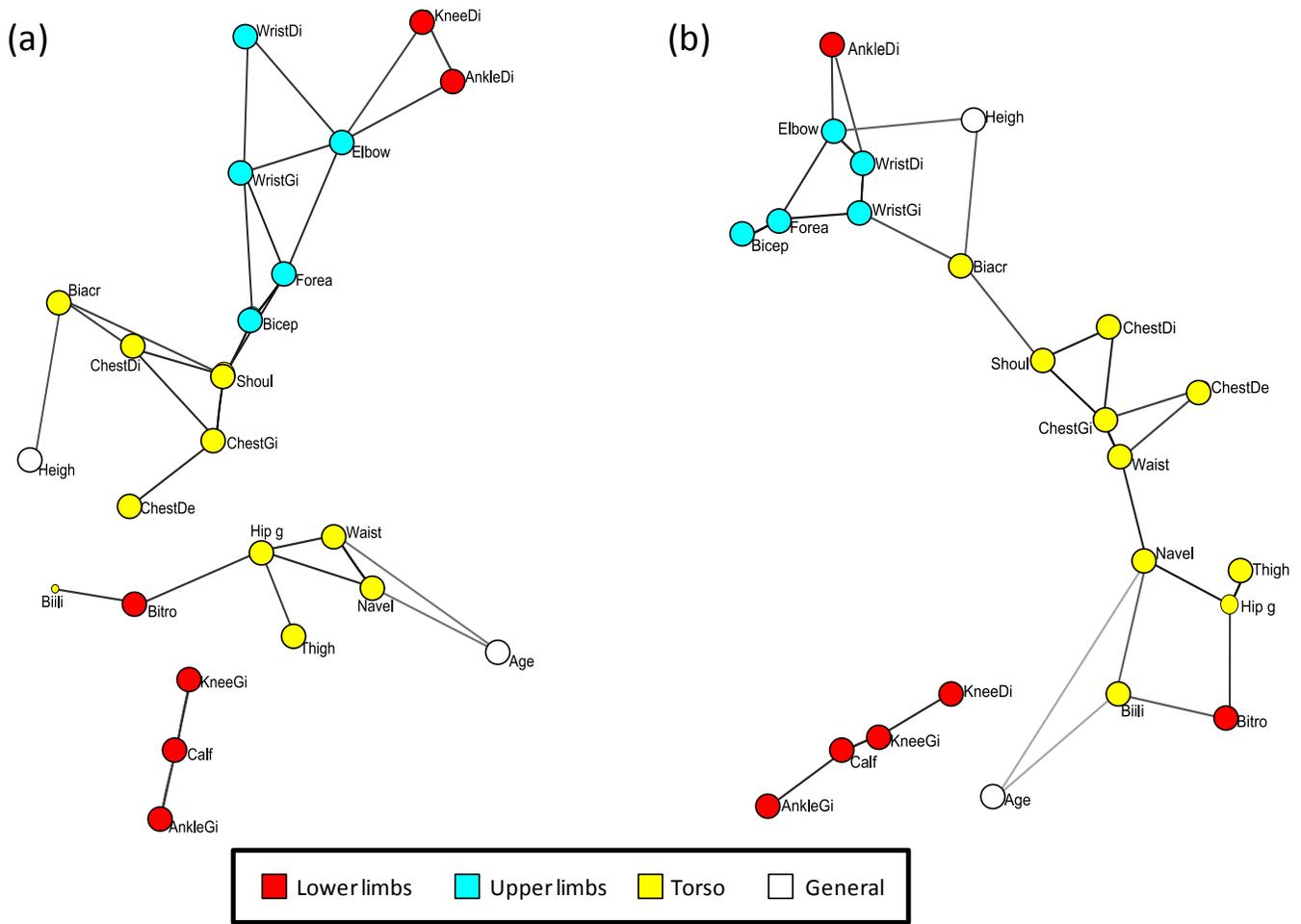


Figure 4. Correlation graphs of the (a) M Group and (b) F Group after the weight vertex and redundant edges were removed by using the MAX2 approach.

could manage to separate males from females based on body shape attributes with an accuracy of 89%. In addition, the role of age on body shape, which is generally a poorly correlated variable when conventional techniques are used, was confirmed with an exceptional high confidence level. The correlation graphs have revealed underlying linear structures in human body shape dynamics with relative ease.

Although relatively confined to date, the potential applications of the correlation graphing methodology appears extensive. This paper demonstrates one of the first applications of correlation graphs in non-time series related data outside an economic context. Also, to the authors' knowledge this is the first paper that presents a correlation graph based on a transposed normalised data set in order to project individual entities in the sample relative to each other. This graph, supported by the

maximum spanning tree revealed exceptional capabilities to surface valuable relationships between entities. Unfortunately the correlation coefficient detects linear trends only and future research may be required to incorporate non-linear trends between variables into the correlation graph. This paper demonstrates how categorical data (such as male vs. female, old versus young) can become part of a correlation graph through respective vertex colouring and subsequent statistical analyses. Correlation graphs have strong capabilities to help sub-divide a system into smaller clusters of elements that are more integrated and uniform in their responses. This kind of analysis, to the authors' knowledge, has not been demonstrated in the literature before and in this paper we have made a concerted effort to present the methodologies in a relatively simple and generalised fashion to stimulate adoption across various

disciplines.

The MAX2 and min- δ techniques are novice, objective and powerful algorithms to remove redundancy and can be used to (a) help quantify the degree of complexity of a system, (b) reveal the dominant factors that regulate the system, and (c) reveal the more subtle factors that may silently in the background play an important role in regulating a system.

In addition to being a powerful multivariate analyses methodology, the visual strength and intuitive understanding of correlation graphs may be valuable when scientists interface with society, especially when relatively complex systems need to be understood. People are generally familiar to visually enhanced maps and trees and intuitively understand graphs that represent relationships between different parts of a system. This does not only create an interfacing opportunity between scientists and society, but also holds potential to be a powerful teaching aid. Research is also needed to further link conventional graph theory descriptive parameters, such as the network's diameter, planarity and degree distribution to properties of the real systems that are under investigation.

REFERENCES

- Baek SY, Lee K (2012). Parametric human body shape modelling framework for human-centred product design. *Computer-Aided Des.* (44):56-67.
- Ballabio D, Vasighi M, Filzmoser P (2013). Effects of supervised self organising maps parameters on classification performance. *Anal. Chim. Acta* 765:45-53.
- Bezuidenhout CN, Van Antwerpen R, Berry SD (2012). An Application of Principal Component Analyses and Correlation Graphs to assess Multivariate Soil Health Properties. *Soil Sci.* Accepted for publication.
- Bulluck LR, Brosius M, Evanylo GK, Ristaino JB (2002). Organic and synthetic fertility amendments influence soil microbial, physical and chemical properties on organic and conventional farms. *Appl. Soil Ecol.* 19:147-160.
- Clark I, Harper W (2008). *Practical Geostatistics 2000*, first ed. Ecosse North America LLC, Columbus, Ohio USA.
- Cuomo F, Cianfrani A, Polverini M, Mangione D (2012). Network pruning for energy saving in the Internet. *Comput. Netw.* 56(10):2355-2367.
- De Nooy W, Mrvar A, Batagelj V (2012). *Exploratory social network analysis with Pajek*.
- Gil MV, Calvo D, Blanco D, Sanchez ME (2008). Assessing the agronomic and environmental effects of the application of cattle manure compost on soil by multivariate methods. *Bioresour. Technol.* (99):5763-5772.
- Gómez S, Jensen P, Arenas A (2009). Analysis of community structure in networks of correlated data. *Phys. Rev.* (80):016-114.
- Heimo T, Tibely G, Saramaki J, Kaski K, Kertesz J (2008). Spectral methods and cluster structure in correlation based networks. *Physica A-stat. Mech. Appl.* 387:5930-5945.
- Heinz G, Peterson LJ, Johnson RW, Kerk CJ (2003). Exploring relationships in body dimensions. *J. Stat. Edu.* 11(2). Internet. Available from: <http://www.amstat.org/publications/jse/v11n2/datasets.heinz.html>. Accessed date 11 July 2013.
- Kamada T, Kawai S (1989). An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* 31:7-15
- Ke Y, Cheng J, Ng W (2012). Efficient Correlation Search from Graph Databases. *IEEE Trans. Knowl. Data Eng.* 20:1601-1615.
- Larsen LG, Choi J, Nungesser MK, Harvey JW (2012). Directional connectivity in hydrology and ecology. *Ecol. Appl.* 22(8):2204-2220.
- Mantegna RN (1999). Hierarchical structure in finance markets. *Eur. Phys. J.* 11:193-197.
- Newman MEJ (2011). *Complex Systems: A Survey*. *Am. J. Phys.* 79:800-810.
- Onnela JP, Chakraborti A, Kaski K, Kertesz J, Kanto A (2003). Asser trees and asset graphs in financial markets. *Phys. Scripta* (106):48-54.
- Onnela JP, Kaski K, Kertesz J (2004). Clustering and information in correlation based financial networks. *Eur. Phys. J.* (38):353-362.
- Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsas D, Anthemidis A, Sofoniou M, Kouimtzi S (2003). Assessment of the surface water quality in Northern Greece. *Water Res.* 37:4119-4124.
- Singh KP, Malik A, Mohan D, Sinha S (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomati River (India) – A case study. *Water Res.* 38:3980-3992.
- Slot M, Bruins HE, Reitsma JB, Rutten FH, Frans H, Hoes AW, van der Heijden GJMG, Geert JMG (2010). Heart-type fatty acid-binding protein in the early diagnosis of acute myocardial infarction: A Systematic Review and Meta-Analysis. *Heart* 24(96):1957-1963.
- Souma W, Maskawa J, Fukuda K (2009). Estimation of network from meaningful part of cross-correlation matrix. *Proceedings of Applications of Physics in Financial Analysis*, Tokyo, Japan.
- Ye R, Wright AL (2010). Multivariate analysis of chemical and microbial properties in histosols as influenced by land-use types. *Soil Till. Res.* 110:94-100.
- Yoshikawa T, Lino T, Lyetomi H (2012). Observation of frustrated correlation structure in a well-developed financial market source. *Progress Theor. Phys. Suppl.* 194:55-63.
- Yuan DG, Mang GL, Gong ZT (2008). Numerical approaches to identification of characteristic soil layers in an urban environment. *Pedosphere* 18:335-343.
- Zhang H, Berg AC, Maire M, Malik J (2006). SVM-KNN: discriminative nearest neighbour classification for visual category recognition. *Proc. of the IEEE Conference On Computer Vision and Pattern Recognition* pp. 2126-2136.
- Zhang JH, Bohme JF, Zeng YJ (2005). A nonlinear adaptive fuzzy approximator technique with its application to prediction of non-stationary EEG dynamics and estimation of single-sweep evoked potentials. *Technol. Health Care* 13:1-21.