UC Davis

UC Davis Previously Published Works

Title

Critiquing the critics: an approach for the comparative evaluation of critical schemas

Permalink

<https://escholarship.org/uc/item/080679wp>

Journal

Social Science Research, 34

Authors

Hsu, Greta
Podolny, Joel

Publication Date

2005

Peer reviewed

Critiquing the critics: an approach for the comparative evaluation of critical schemas[☆]

Greta Hsu^a and Joel M. Podolny^{b,*}

^a *Graduate School of Business, Stanford University, Stanford, CA 94305, USA*

^b *Department of Sociology, Harvard University, Cambridge, MA 02138, USA*

Available online 28 March 2004

Abstract

The emergence and persistence of a market requires a broadly accepted framework or schema for the evaluation of product offerings. In many markets, critics play an important role in the elaboration of these schemas. However, sociologists lack a systematic method for comparing critical schemas and consequently lack a systematic method for assessing how critics play their respective roles. In this paper, we propose a method for making such comparisons. This approach gives rise to a role typology, where schemas can be of four types: simple, complex, unpartitioned, and robust. This paper applies this method to the analysis of film reviews appearing in *Variety* and the *New York Times* between 1946 and 1982.

© 2004 Published by Elsevier Inc.

Keywords: Roles; Markets; Schemas; Critics

---

1. Introduction

A necessary condition for the emergence and persistence of a market is the existence of a broadly accepted framework or schema that enables consumers to compare producer offerings. Such an observation has its origins in White's (1981) conception of a market as a role structure. In White's model, the actual volume-revenue choices of market producers yield an orderly array that provides consumers and

---

[☆] We wish to thank Jim Baron, Glenn Carroll, Mark Granovetter, Mike Hannan, Miller McPherson, and Ezra Zuckerman for helpful comments on this paper. Earlier versions of this paper were presented at the Stanford 2001 Strategy Conference and the Harvard Department of Sociology.

* Corresponding author.

E-mail addresses: sprocket@stanford.edu (G. Hsu), jpodolny@wjh.harvard.edu (J.M. Podolny).

producers with a tangible guide as to whether or not potential alternative terms of trade would represent a “fair value” that the consumers should and therefore would accept. Such an observation also underlies the status-based model of market competition (Podolny, 1993); the existence of a status ordering induces a differentiated yet stable flow of payments from consumers that, in turn, leads producers to make investment decisions and pursue exchange relations that reproduce the ordering.

Probably the most provocative demonstration of the impact of a framework or schema on market outcomes is Zuckerman’s (1999, 2000); examination of capital markets. The work is notable for at least two reasons. First, Zuckerman documents a penalty that a firm pays when its portfolio of assets or activities violates the framework or schema on which market decisions are based. So, in his particular case, Zuckerman shows that a firm experiences a discount to its stock price when its portfolio of assets cuts across the cognitive categories established by financial analysts for sorting firms into relatively well-defined reference groups. The second important feature of Zuckerman’s work is that it draws attention to the role of third parties—analysts, or perhaps more generally, critics—as a source of a framework or schema. In this respect, Zuckerman’s research resonates with studies of cultural markets that explore the role of these third parties in establishing and legitimizing a framework that guides the evaluations of others (Becker, 1990; Greenfeld, 1989; White and White, 1992).

Yet while sociological researchers have identified the important role played by critics, they have not developed a metric or a methodology for evaluating *how* a critic plays his or her role. Put somewhat pithily, there exists no analytical basis for critiquing the critic. Given the initial observation that a well-defined evaluative schema is a necessary precondition for a functioning market, it seems important to develop such an analytical basis for evaluating schemas. Once such a methodology is developed, it should be possible to relate features of the underlying framework to a variety of market outcomes. The next section develops this methodology, first forming the basic ideas for comparing quantitative one-dimensional schemas and then extending these ideas to contexts in which critics employ multidimensional schemas for evaluating market products. The model is then applied empirically to the critical reviews of US feature films from 1946 to 1982.

2. A method for evaluation

Research on cultural markets is probably the most explicit in defining the role of the critic in terms of the development of a framework or schema for evaluation. For example, in *Careers and Creativity* White (1993) refers to critics as “ex-post tellers of tales” (p. 57) who “manipulate sets of narratives into a new creative narrative” (p. 59). They do so primarily by “partitioning by archetypes, by phases, by modes, and by themes (p. 60).”

Viewed in isolation, a critic’s review is simply a judgment about qualities of an individual act or work. However, reviews do not exist in isolation. From the vantage of the critic, the release of artistic works are the raw materials and the opportunities

from which the critic creates his or her own edifice, and partitioning through the use of language is the fundamental tool at the critic's disposal. From the vantage of the artist (or more generally, the individual whose effort is evaluated by the critic), the partitioned schema is a sorting mechanism through which his or her effort passes on the way to being evaluated by the broader audience. The slot into which the critic places the work strongly shapes the expectations, perception, and—at a more basic level—the attention of that broader audience. In some instances, the critical edifice may be the construct of a single individual; in others, it may have more of the character of a social movement, with different individuals making partial contributions to a broader whole.

While it is perhaps easiest to perceive the critical function in these terms when contemplating cultural markets, it is not necessary to restrict the discussion to such a domain. For example, consider Hirsch's (1986) account of the legitimization of hostile takeovers. Hirsch observes how the initial opposition to hostile takeovers was transformed into acceptance in part through the use of such language as "golden parachutes," "white knights," and—to highlight a term that denoted a mechanism for thwarting a hostile takeover—"poison pills." As colorful new phrases and descriptors became the lens through which hostile takeovers became viewed, analysts, executives, and investors acquired a much more favorable disposition to this activity. In this instance, the critical function was played largely by journalists who described the takeover activity, but the journalists obviously drew on bankers and executives themselves for much of their terminology. Similar analyses could be applied to the rejection of unrelated diversification.

Journalists, bankers, and academics—all playing a critical function—attached a negative cast to such activity, and this negative cast represented the lens through which investors, the "consumers" of the capital markets, began to see the phenomenon of unrelated diversification (Davis et al., 1994).

Or, to shift to a market in which the roles of consumer, producers, and critic are easier to define, consider the market for wines. In this domain, scholars have been generally sensitive to the role that critics must play in identifying attributes or types that form the basic language by which consumers can discriminate between wines. By application of such terms as "dry," "sweet," "smooth," "complex," "oak-y," and by comparison of wine tastes to different types of fruits, consumers have a taxonomy that sensitizes them to the distinctions between different wines (Adams, 1990).

If the critical function is fundamentally an act of partitioning, then how does one compare different partitions? How does one compare the different critical schemas that are developed? This is the task to which we turn. We begin by considering a simple example.

Assume that a critic uses numbers rather than words to evaluate products. The numbers are arrayed on some scale with a fixed range, say from 1 to 100, and a higher number is considered a superior rating to a lower number. Finally, assume that the critic assigns only one number to each product that appears on the market. So one product might receive a "23" while another might receive a "68." According to this scale, the product that receives the 68 is better than the 23. One can clearly find

examples of critics who employ such a schema to provide their final and objective evaluations of products on a market. For example, *Car and Driver* often rates automobiles in a given class on a 1- to 10-point scale rounded to the nearest 10th of a point, and Robert Parker—perhaps the most prominent wine critic in the US—evaluates wines on a 100-point scale in his newsletter. In both instances, there are typically verbal descriptions that accompany the final number, and sometimes the final number is a sum or average of ratings along a number of dimensions, but for the sake of an initial illustration we will assume a critic that employs only one number and no verbal accompaniment to that number.

We know that over time, the critic will construct a one-dimensional schema due to the restriction that the critic only employs one number for each product on the market. However, these one-dimensional schemas can vary in some important and systematic respects, as illustrated in Fig. 1. There are five horizontal lines in Fig. 1; each represents a continuum from 1 to 100, and there are 10 vertical hash marks on each continuum. Think of each of the hash marks as denoting a particular review. Each of the five distributions of hash marks denotes one of the obviously infinite variety of ways that 10 reviews could be distributed along the continuum.

Distribution A denotes a schema that can be characterized as simple, but intense. The critic's opinions of the products that he or she is evaluating are either strongly positive or strongly negative. Hence, we use the term “intense” to describe the distribution. However, given that the critic is positive about a product, there is little variance in the extent to which he or she is positive; similarly, given that the critic is negative, there is little variance in the extent to which he or she is negative. Essentially, there is little more than a “thumbs-up” or a “thumbs-down” in the evaluation.

Fig. 1. Hypothetical evaluative distributions.

Despite the fact that there are 10 evaluations, there are only 2 clusters or partitions of evaluations.

Distribution B differs from distribution A in its intensity, but not in its simplicity. There is little variance in rating among those products that are evaluated negatively, and there is little variance among those products that are evaluated positively. The only difference between distributions A and B is in the variance between positive and negative evaluations. Stated somewhat differently, once again, there are only 2 clusters, but the distance between those clusters is less for B than for A.

As with distribution A and distribution B, distribution C still represents a clustering or partitioning of evaluations across the distribution. However, in this case, the number of partitions is five rather than two. A simple “thumbs-up” and “thumbs-down” distinction would not capture the nuances in the partitioning of distribution C. For this reason, we label distribution C as complex.

Distribution D reveals little if any evidence of categories or types. While the fourth and fifth hash mark on the left are close together, the nearness of the two hash marks is as likely due to random chance as to the systematic construction of a single type. If we were to try to cluster or partition the D distribution in a way that reasonably reflected the distribution of hash marks, we would need almost as many partitions as hash marks due to the disorderly distribution of marks on the distribution. For this reason, we label distribution D as “unpartitioned.”

Finally, consider distribution E. Like distribution A, distribution E has two reasonably clear partitions at the ends of the continuum. That is, in distribution E, five of the evaluations are reasonably close to the left end of the continuum, and five are reasonably close to the right end of the continuum. However, within these two partitions are clear sub-partitions. Like distribution C, one can find a reasonably high number of pairings (4 pairings in E as compared to 5 pairings in C). Following from the terminology employed by Leifer (1988) and Padgett and Ansell (1993), we use the terms “robust” and “multi-vocal” to describe the distribution in E. For a consumer that wants a “thumbs-up”/“thumbs-down” characterization, a critic yielding a distribution like E provides some reasonably clear guidance—almost as clear as the guidance in A. However, for a consumer wanting a more nuanced, complex evaluation of the products on the market, the schema in E provides almost as much nuance as C. Because the E distribution is of value to the consumer that wants a simple schema and the consumer that desires a more nuanced schema, the term robust or multi-vocal seems a relatively apt description.

There is perhaps a tendency to look at E and think of it as being somehow superior to the other schemas. In our view, such an interpretation of the distinctions between the schemas would be a mistake. Our objective with this stylized example is simply to show that there are multiple ways that a critic can play out his or her role. With the exception of D—where there is no evidence of partitioning—all of the other distributions represent reasonable ways that the critic could enact his or her role. A subsequent—and indeed very important question—is whether different schemas have different implications for producers in the market. For example, do more complex or robust schemas create more niches and in so doing lower the intensity of competition among producers in a market?

minimized. The asterisks above distribution A in Fig. 2 represent the locations that would indeed minimize this summation. Each point can be seen as being at the center of its own cluster, which is composed of those evaluations that are closer to it than to the other ‘*.’ Thus, we can label the ‘*’ symbols as cluster center-points.

If a distribution is a simple thumbs-up/thumbs-down distribution, then a two center-point approximation to the over-all distribution should closely resemble the over-all distribution, and in distribution A, this clearly seems to be the case. However, how do we formally assess how well the two center-point approximation represents the over-all distribution?

First define the mean evaluation in the distribution

$$\bar{e} \equiv \sum_{i=1}^{10} e_i / n.$$

Given this definition of the mean evaluation, the value

$$\sum_{i=1}^{10} (e_i - \bar{e})^2,$$

is analogous to the total sum of squares in a regression context.

Similarly, the expression

$$\min \left[\sum_{i=1}^{10} \min_{k=1,2} (e_i - p_k)^2 \right],$$

which is the sum of within-cluster distances when the points p_k are chosen to best reflect the underlying distribution, is analogous to the residual sum of squares in a regression context. With the clear analogues to total sum of squares and residual sum of squares, the expression

$$\frac{\min \left[\sum_{i=1}^{10} \min_{k=1,2} (e_i - p_k)^2 \right]}{\sum_{i=1}^{10} (e_i - \bar{e})^2},$$

is the proportion of variance that is unexplained by the two-point distribution.

While we use the two-point distribution for the purpose of illustration, it should be clear that the notion of unexplained variance by the two-point distribution generalizes to m points (i.e., cluster centers) where m is less than or equal to the n evaluations in the distribution. For m cluster centers and n evaluations, the proportion of variance unexplained would be represented as

$$\rho_m \equiv \frac{\min \left[\sum_{i=1}^n \min_{k=1,m} (e_i - p_k)^2 \right]}{\sum_{i=1}^n (e_i - \bar{e})^2}.$$

We label this value ρ_m because it can be interpreted like the intraclass correlation coefficient ρ in random effects estimators (Tuma and Hannan (1984), pp. 438–442). The expression ρ_m represents the proportion of the total variance that is within clusters and hence not captured or explained with the location of the cluster centers when

there are m clusters. Conversely, the value $1 - \rho_m$ represents the total proportion that is explained by the distribution.

Fig. 2 depicts the central points of the 2-cluster and 4-cluster solutions that minimize the unexplained variance in each distribution in Fig. 1. The center-points of the 2-cluster distributions are represented by the '*' symbols; the center points of the 4-cluster distributions are represented by the '+' symbols.

Comparing across distributions, it is noteworthy that they differ considerably in the extent to which 2- and 4-cluster solutions reduce the proportion of unexplained variance in the overall distribution of evaluations. As noted above, the 2-cluster solution explains a considerable proportion of the variance in the simple distributions (A and B), especially when compared to the complex distribution (distribution C). However, with the simple distribution, 4 clusters do not explain much more of the variance in the overall distribution of the 10 evaluations than two points; in contrast, with the complex distribution, the additional two points considerably increase the explained variance.

Neither the 2-cluster or 4-cluster solution explain much of the variance in the unpartitioned distribution of evaluations (distribution D); one would need to use 6 or 7 clusters before one explained as much variance as the 2- or 4-clusters explain in distribution A through C. Finally, consider the robust or multi-vocal distribution. The center points of the 2-cluster solution are a slightly worse representation of a robust or multi-vocal distribution than of the simple distribution, but two points better represent the robust distribution than the complex or unpartitioned distribution. However, whereas the additional 2 clusters do not improve the variance of the simple distribution that is explained, the additional clusters do increase the explained variance of the robust distribution.

With these distinctions, we can now introduce the idea of a cluster profile. Fig. 3 provides an illustration of how additional clusters reduce variance in the distributions. The simple and robust distributions experience an initial rapid drop in proportion of variance unexplained. The complex and unpartitioned do not. The complex distribution reveals a constant reduction in variance across the entire distribution; indeed, one might expect a complex distribution to be associated with a steepening of the curve once there are enough distinct points to reflect the numerous partitions in the data. With the unpartitioned distribution, a considerable percentage of the variance remains unexplained. In effect, the different distributions can now be represented with different profiles.

Alternatively, one can think about representing these differences in a two-dimensional plane; the vertical axis represents the variance explained where m , the number of clusters is low, and the horizontal axis represents the additional variance explained when m is increased. Given such a two-dimensional plane, each quadrant will be associated with a particular profile, as represented in Fig. 4.

Adjacent quadrants obviously share a common attribute. For example, simple and robust schemas both explain a reasonable amount of variance with low m , and simple and unpartitioned schemas are alike in so far as additional clusters do not add significantly to the variance explained. However, the schema denoted by a particular quadrant also shares a common attribute with the quadrant in the

Fig. 3. Cluster profiles.

Fig. 4. Two-dimensional representation of evaluative schema.

opposite corner. Simple and complex schemas are alike in that there is comparatively little unresolved ambiguity in the meaning of the partitioning. A simple schema is one in which products are clearly sorted into few types; a complex schema is one in which

products are clearly sorted into many types. In contrast, neither the unpartitioned nor robust schemas lend themselves to unequivocal sorting of product offerings.

In systematically comparing the differences between schemas, we have purposely avoided consideration of the moments of the distribution. The reason is that we needed to elaborate the concept of the cluster profile and the two-dimensional trade-off plane before we could consider the substantive implications of differences in the key distributional moments. We will focus on what seem to us to be the two most important—variance, which is the second moment, and skew, which is the third moment.

Greater variance implies greater intensity; as just discussed, the difference between a high-variance and low-variance distribution is reflected vividly in the distinction between the two simple distributions—A and B—in Figs. 1 and 2. Notably, when one represents the more intense distribution with either two or four points, one captures a greater proportion of the underlying variation. More generally, holding all other features of a distribution constant, the greater the variation among a set of evaluations, the more variation is “explained” with a given number of clusters. As a consequence, the higher the variance between evaluations in a distribution, the more that the distribution will tend to locate away from the bottom-left corner within the unpartitioned quadrant.¹

One can reasonably ask whether such a shift is phenomenologically accurate. Does it make seem sensical that intensity implies greater partitioning—be that partitioning simple, robust, or complex? In our view, it does seem appropriate that greater intensity or variance would imply heightened linguistic structuring. Enhanced variance is one way in which a critic can establish distinctions between observations. However, suppose that one disagrees and regards intensity as an orthogonal attribute of a linguistic schema that should not affect a schema’s location in the two-dimensional plane in Fig. 4. In this case, one could normalize the distances in a set of schemas such that the average distance of an observation from the mean is the same for all schemas. More formally, in the above expressions, one replaces

$$(e_i - \bar{e}),$$

which is the average distance of each evaluation from the mean, with the expression

$$\frac{(e_i - \bar{e})}{\sum_{i=1}^n (e_i - \bar{e})/n}.$$

In this way, one can effectively eliminate differences in the variance across schemas if one chooses.²

¹ One cannot specify whether higher-variance implies more of an upward vertical shift or a rightward horizontal shift; the particularities of the shift will depend on other features of the distribution. Therefore, it is more accurate to say that higher variance implies a location away from the lower-left corner rather than toward any other particular corner.

² In the empirical analysis that we conduct later in this paper, we compared schemas with and without this type of normalization. At least in this particular analysis, the normalization had little practical consequence for the location of schemas in the two-dimensional space.

Now consider the third moment, the skew of the distribution. With the exception of the unpartitioned schema, each of the evaluative distributions in Figs. 1 and 2 is evenly balanced. However, suppose, for example, that a critic gave nine films a “thumbs-up” and only one film a “thumbs-down.” In this instance, it should be clear such asymmetry would cause the evaluative distribution to appear extremely simple. One does not need even a 2-cluster solution to explain more than 50% of the variance in the underlying distribution. One can reproduce 90% of the distribution with only one center-point. The same trend will apply generally. The more that a critic assigns a disproportionately large number of observations to a small number of categories, the easier it will be to represent the mass of the distribution with a small number of cluster centers. In our view, the impact of skew on the location of distribution in the two-dimensional plane is—like the implication of higher variance—substantively appropriate; the disproportionate use of a small number of categories belies a lack of subtlety and complexity.

We have now completed our illustration of how one might formally compare different evaluative distributions along certain key dimensions. In the next section, we move to the context that will be the focus of our empirical analysis—film reviews. For a film critic, each new film is an object that provides the critic with the opportunity to build and reinforce his or her particular schema, and words are the tools by which the schema is established. In the highly stylized example of Fig. 1, the schema was simply a numerical distribution that was generated by the distances between evaluations. In order to be able to quantify features of a film critic’s schema, we need a way to measure the “distance” between verbal evaluations just like we can measure the distance between evaluations on a 100-point scale. For example, suppose that a film critic uses the terms “excellent,” “extraordinary,” and “wonderful” to describe film 1 and uses the terms “satisfactory,” “competent,” and “expressive” to describe film 2, we need a way of assessing the distance between those evaluations.

One way to do so is to rely on what can perhaps best be called synonym overlap. One takes the words used to describe film 1 and all of the synonyms of those words (i.e., the synonyms of excellent, extraordinary, and wonderful). Then, one takes the words used to describe film 2 and all of the synonyms of those words (i.e., all of the synonyms of satisfactory, competent, and expressive). One then sees whether there is any overlap across the two groups of words. One could obviously construct a proportional overlap measure. For example, if 50% of the union of all words employed to describe either of the two films are used to describe both films, then one could assign a synonym overlap value of .5. If the identical words and accordingly identical synonyms are employed, then the value would be 1. Conversely, the distance between two films is 1 minus the proportion of overlap.³

---

³ There are other distance measures that one could use. For example, one problem with the synonym overlap is that the measure does not strongly distinguish word pairs that are simply not synonyms from word pairs that are antonyms. For example, the distance between “fast-paced” and “reasonable” may be identical to the distance between “fast-paced” and “slow-paced.” One could perhaps develop a path distance measure in a word matrix to pick up such distinctions. However, since the purpose of this paper is simply illustration, we will focus on only one measure of distance, synonym overlap.

If a critic reviews N films, one can represent the synonym overlap across the films with an $N \times N$ matrix, where cell i, j denotes the synonym overlap of films i and j . One then can apply the partitioning to the rows of the matrix. Two films, i and j , have a similar position in the critical landscape if row i and j of the matrix are similar (i.e., if i and j have the a similar pattern of synonym overlap with all of the other films). Conversely, two films, i and j , occupy very different positions in the critical landscape if the values in row i are very different from the corresponding values in row j . So, by treating each row as an observation and applying a clustering method to the data, it is possible to derive cluster profiles like those depicted in Fig. 3, and it is possible to position different critical landscapes in the two-dimensional place represented in Fig. 4.

It is important to note that the choice of an appropriate method for cluster analysis is a crucial aspect of this process. Clustering methods differ in a number of respects, and markedly different results can emerge when different methods are applied to the same set of data. When choosing a clustering method, it is important to consider the specific goals of the analysis and how these will be best met by the particular ways in which a method identifies the “best” partition for a set of data.

One of the most basic distinctions between methods is that which exists between those methods that develop a hierarchically nested set of partitions and those methods that apply a single partition for a fixed number of clusters. The first type, referred to as hierarchical clustering, yields an “evolutionary tree” through a series of path dependent partitions. Hierarchical algorithms proceed in either an agglomerative or divisive fashion. Agglomerative methods begin by treating each data point as a singleton cluster and iteratively merge the most similar pair of clusters. Divisive algorithms, by contrast, begin with a single cluster and determine, at each step, the best division possible among extant clusters. Differences between methods arise as a result of the different criteria used to define similarity. While hierarchical methods are very popular, the stepwise optimality of hierarchical methods can be problematic for researchers since algorithms implementing these methods are unable to repair erroneous decisions made during previous steps (Kaufman and Rousseauw, 1990). Partitioning methods, by contrast, directly seek the best partition for a set number of clusters. Algorithms implementing these methods generally specify an initial partition of data points into the specified number of clusters and then iteratively modify this partition until none of the allowable transformations will result in an improvement in the clustering criterion. Partitioning methods are likely to explore a greater range of partitions for a given number of clusters compared to their hierarchical counterparts and are therefore more appropriate for comparisons of the explanatory power of different numbers of clusters.

Different partitioning methods can be distinguished from one another by the numerical criterion used to assess the adequacy of partitions. The basic goal of partitioning is to produce clusters that are internally cohesive and well isolated from one another. Measures of partitioning adequacy thus focus on minimizing heterogeneity within clusters, maximizing separation between clusters, or are based on a

combination of the two goals.⁴ Examples of measures of the internal heterogeneity of a cluster are the dissimilarity between the most dissimilar pair of points in a cluster, the sum of the distances between the median point and all other points in the cluster, and the sum of distances between every pair of points in the cluster. Example measures of the isolation of a cluster are the smallest dissimilarity between any point in a cluster and any point outside of that cluster and the sum of the dissimilarities between every points in the cluster and every point outside the cluster.

One of the most commonly used criteria for partitioning is minimization of the sum of squared Euclidean distances between every point in a cluster and their group mean. For this criterion, minimizing heterogeneity within clusters is equivalent to maximizing isolation among clusters. That is, by minimizing within-cluster sum of squares, this method at the same time maximizes between-cluster sum of squared distances. Methods employing this criterion offer a particularly direct means for assessing cluster profiles since the variance explained by each partition is simply the ratio of between-cluster sum of squares to total sum of squared distances. Because of its intuitive simplicity, it is the method we choose to employ in our analyses.

MacQueen's (1967) *k*-means algorithm is a popular clustering algorithm that implements the sum of squares criterion when constructing clusters. Like other partitioning algorithms, *k*-means begins with an initial partition of the data into the specified number of clusters. The partition is iteratively updated by relocating each data point to the cluster with the nearest group mean and recalculating group means. The algorithm runs until no change in cluster membership will improve the clustering criterion.

While initial implementations of the *k*-means algorithm assigned objects to clusters randomly, Monte Carlo studies of the performance of iterative methods have shown that using results from hierarchical clustering to inform the initial starting position for *k*-means can result in superior recovery of data structures when compared to the performance of other partitioning and hierarchical methods (Milligan, 1980). Following this, we use the results of Ward's (1963) hierarchical clustering algorithm to provide initial starting seeds for *k*-means clustering. Ward's algorithm is chosen because it also utilizes the sum of squares criterion when choosing partitions—at each step, its objective is to choose the fusion of clusters that minimizes increase in total within-cluster sum of squares.

It is important to emphasize that what we are proposing is a general approach to the problem of developing cluster profiles—one which should be tailored to the specific needs of the data under observation. The choice of a specific clustering criteria will invariably impose various assumptions about the data's structure on empirical findings. Therefore, the nature of the data and anticipated clusters are important to consider when choosing among clustering methods since use of an inappropriate method of analysis may result in distorted recovery of the real structure of the data. For example, use of the sum of squares criterion has empirically been observed to produce clusters with roughly equal numbers of objects which occupy hyperspherical regions in space (Gordon, 1999). Our use of *k*-means with Ward's seeds is therefore

---

⁴ For reviews of partitioning criteria, see Everitt et al. (2001); Gordon (1999).

most appropriate when clusters are suspected to be roughly equal in size and spherical in nature. When clusters are suspected to be non-spherical or unequal in size, single-linkage or algorithms implementing statistical mixture models (Banfield and Raferty, 1993; Scott and Symons, 1971) may be more appropriate.

Fig. 5. Stylized matrices.

In many situations, it is not possible to determine a priori a single preferred clustering method. It is generally useful to analyze the same data using two or more methods. If the results of the different analyses are similar, one can have greater confidence in the empirical results. In our own analyses, we first analyze and present the results of using k -means with seeds from Ward's to demonstrate the formal comparison of different schemas for the evaluation of films. We then compare these results to those obtained through another partitioning method in an effort to assess the reliability of these results.

However, before moving to the empirical analysis, it is useful to illustrate how different stylized patterns of overlap in a $N \times N$ matrix are manifest in the two-dimensional plane of Fig. 4. So far, we have only considered one-dimensional distributions of evaluations, and the addition of a dimension brings added complexity to the critical landscapes. By considering some ideal-typical matrices, we can hopefully give the reader an intuitive sense for what causes an overlap pattern to be simple, complex, robust, or unpartitioned.

Consider the eight different patterns represented in Fig. 5. The first of these patterns constitutes two separate clusters of unequal size. Such a pattern would arise if a critic, for example, simply labeled some films as "good" and others as "bad" but was not evenly predisposed to be positive or negative. The second of the patterns represents four separate clusters of equal size. Such a pattern indicates a richer linguistic landscape than "good" or "bad," but not much richer. For example, such a pattern would arise if a critic labeled films as either "funny," "dramatic," "long," or "shallow" and sought to ensure that an equal number of films were allocated to each category. The third matrix is like the second except there is a fifth separate cluster, and the fourth matrix represent five separate clusters with one of the clusters being notably larger than the rest.

In the first four matrices, each cluster is completely separate from the rest, but this is not the case in the fifth matrix. We label the fifth matrix as bimodal because it has two "centers." A critic solely employing the descriptors "pretty good," "good," "very good," "pretty bad," "bad," and "very bad" would yield a pattern like this one. The words "good" and "bad" constitute the centers of the bimodal distribution. The sixth matrix reflects a normal distribution, where there is a comparatively large mass of synonyms in the center of the distribution that tails off into a few extreme judgements on both ends. The seventh matrix is a completely random pattern of overlaps across films; such a pattern suggests no systematic attempt to partition. Finally, the eighth matrix depicts a uniform distribution; each film has a set of descriptors that imply an overlap with a fixed number of linguistically adjacent films. In effect, this implies a "chain" landscape.

While the matrices in Fig. 5 are 10×10 , we constructed 50×50 matrices of each stylized type for the purpose of analysis. The reason is simple; we wanted N to be considerably larger than the number of clusters used to represent the underlying distribution.⁵

⁵ In going from a 10×10 matrix to a 50×50 matrix, we essentially allow each cell in the 10×10 to be represented by a 5×5 combination of cells. The only exceptions are the normal distribution and bimodal distribution, which require additional smoothing of the pattern in going from a 10×10 to a 50×50 , and the random distribution, which would not be random if a cell's value was determined by an adjacent cell.

Fig. 6 depicts the position of the matrices in the two-dimensional space employing various clustering algorithms. We compare hierarchical agglomerative algorithms (single- and average-linkage and Ward's) as well as partitioning algorithms (k -means with initial seeds from Ward's and Partitioning Around Medoids (PAM)). As stated earlier, the agglomerative hierarchical methods can be distinguished by the criterion used to assess similarity between clusters. Under single-linkage (McQuitty, 1957; Sneath, 1957), the similarity between clusters is measured by the distance between the closest pair of objects from each group. Average-linkage (Sokal and Michener, 1958), by contrast, measures similarity as the average distance between every pair of objects between each group. And Ward's method measures similarity between each pair of clusters by the increase in error sum of squares that would result from the merging of those 2 clusters. The two partitioning methods of k -means and PAM similarly differ in the criterion they choose to optimize. While k -means minimizes within-cluster sum of squared distances, PAM minimizes the sum of distances of each object to its cluster's medoid (also termed the centroid, or representative object in its cluster).

In this figure, the vertical axis represents the percent of variance explained by a 2-cluster representation of the data; the horizontal axis represents the increment explained by the 6-cluster representation. The dotted line denotes a 'feasibility frontier'; at this line, the percent of variance explained by the 2-cluster solution and the increment to the percent variance explained from the 6-cluster solution equal 100. Thus, it is not possible for observations to be to the right of this frontier line. All of those patterns that had six or fewer separate clusters necessarily line up on this frontier

Fig. 6. Distribution of pure types using different clustering methods.

line. Instead of drawing such a frontier, one could transform the horizontal axis so that it reflects the percent of variance that can be possibly explained.⁶ However, if the empirical analysis in the following section is any indication, “real” matrices—drawing on the actual reviews of critics—are much further from the frontier than these ideal-types. Therefore, there is little reason to introduce the added complexity into the definition of the horizontal axis.

We should note that when different clustering algorithms yield an identical position for a given matrix, we adjust the locations in the figure slightly so that it is easier for the reader to perceive and interpret the results. For example, all of the clustering algorithms yield an identical location for the two-clique unequal matrix ($\rho_2 = 100$, $\rho_6 = 0$). However, the representation is adjusted slightly.

There is a obviously a certain arbitrariness in the placement of the dividing lines separating the quadrants. Complexity, simplicity, robustness, and a lack of partitioning are relative rather than absolute judgments. Because of this arbitrariness, we have drawn the quadrants so that an equal number of observations appear on either side of the horizontal and vertical dividing lines, but other decision rules are possible. For example, one could use the mean or median value on each axis to divide the quadrants.

Regardless of the actual placement of the lines, a comparison of the location of the matrices should provide the reader with an appreciation for how different aspects of the overlap pattern affect the location of a schema in this two-dimensional space. As Fig. 6 shows, there is general agreement among the different clustering algorithms regarding the position each stylized type is assigned to in the two-dimensional space. Noticeably, single-linkage differs from the other methods in regards to where the normal, uniform, and random distributions belong in this space. This is probably because these stylized distributions do not have naturally well-defined groups of objects, and when clusters are not well-separated, single-linkage has a tendency to produce longer “chains” of objects compared to other methods.

The simplest schema is obviously the 2-cluster matrix since it is possible to completely reflect the underlying distribution of the observations with the 2-cluster solution. After the 2-cluster matrix, the next simplest distribution is the 5-cluster matrix where 1 cluster is of much greater size than the others. Compare the location of this matrix to the location of the 5-cluster matrix where the clusters are all of equal size. When the clusters are of equal size, the 5-cluster matrix appears complex. This comparison illustrates the earlier observation as to how the degree of skew in the cluster distribution will impact on the location of a schema. If a comparatively large number of observations becomes grouped into 1 cluster, the schema is appropriately classified as simpler since it fails to draw distinctions between a comparatively large number of observations.

As one would expect, the random pattern is located the furthest into the unpartitioned quadrant. The bi-modal matrix is the one that is most clearly in the robust

⁶ More specifically, one could divide the percent improvement from 6 clusters by the quantity: $100\% - \rho_2$.

quadrant. The 2-cluster solution represents a reasonable approximation of the underlying data in the bi-modal matrix since the six categories can be aggregated into two broad groups. At the same time, the 6-cluster solution picks up the remaining variance. Because of the location of the horizontal line, both the bimodal and 4-cluster matrices appear in the robust quadrant. However, the relative position of the two is noteworthy. The 4-cluster matrix lies closer to the border of the complex quadrant because the clear divisions between its clusters ensures that it cannot be as easily represented with a 2-cluster solution.

Another illustrative comparison is between the matrix with the uniform distribution and the matrix with the normal distribution. One should not draw a particularly strong inference from the fact that one of the matrices is to the left of the central vertical line, and one of the matrices is to the right. As noted above, there is a certain arbitrariness to the location of the borders between quadrants. Both distributions lend themselves to clearer partitioning than the random matrix, and both are less complex than the equally sized 5-cluster matrix. They are thus intermediate or hybrid types. However, the relative position of these two distributions is worth reflecting upon. The normal distribution is higher and further to the left because there is a “center” to the normal distribution allowing for a greater mass of the distribution to be represented with a small number of clusters and, at the same time, implying less improvement in representation with additional clusters. But, what is most noteworthy is that the normal distribution is closer to the lower-left corner than the uniform distribution. This fact is important because one way to think about the difference between the normal and uniform distribution is the central mass of the normal distribution becomes “pushed out” into the tails of the uniform distribution. Stated differently, the uniform distribution is like the normal distribution in the first and third moment, but differs from the normal distribution in the second moment. This difference in locations thus illustrates the point above: higher variance pushes a distribution out from the lower-left corner.

We have used the stylized matrices to help provide the reader with an intuitive sense for what attributes of a two-dimensional distribution cause the distribution to be categorized as simple, complex, robust, or unpartitioned. With this background, we now shift to an examination of the critical film reviews.

3. Empirical context: hollywood films

3.1. Data collection

The present study focuses on the reviews of feature films distributed in the US between 1946 and 1982. Films released in every third year during this period were included in the study sample.⁷ To build a list of films, movie titles were extracted from

---

⁷ The sample excludes documentary and non-US produced films.

Fig. 7. Count of films by data source.

Microsoft's *Cinemanía*'94, an electronic database that lists movie credits and other descriptive information for films distributed up until 1992.

Critical reviews for films listed in *Cinemanía* were then gathered from the *New York Times* and *Variety*. These two periodicals hold several attractions for gathering data on critical schemas. Both reviewed films for the entire period of study, and both covered a relatively large proportion of movies listed in *Cinemanía*.⁸ Fig. 7 shows the number of films listed in *Cinemanía* and reviewed in the *New York Times* and *Variety* from 1946 to 1982.

Additionally, differences in the primary audiences of each periodical are likely to result in interesting and important differences in the evaluative distributions of critics for each. The *New York Times* is a nationally circulated daily periodical geared towards the mass audience of movie-goers. *Variety*, on the other hand, is a weekly trade journal whose primary audience is professionals in the entertainment business. Quite possibly, how a critic perceives and understands her role will vary according to the interests and needs of her primary audience.

To generate measures of synonym overlap, an extensive list of synonyms was obtained from Webster's online thesaurus (version 1.5), an electronic database of words and their synonyms.

⁸ The *New York Times* reviewed 64.59% of *Cinemanía* films while *Variety* reviewed 83.57%. One thousand four hundred and thirty-four films (60.2% of total) were critically reviewed by both periodicals.

3.2. Constructing film-by-film matrices

For every film review in the sample, each adjective used by the reviewer was recorded and assigned a code depending on the aspect, or dimension, of the film it was in reference to. The different dimensions included in the analyses are plot, screenplay, and subject matter. Adjectives belonging to the plot dimension refer to the film's story line. Examples of plot adjectives would be an "exciting" plot or "confusing" story. Adjectives in the screenplay category are in reference to the quality or type of script (e.g. "trite" dialogue or "complicated" wording). And the dimension of subject matter refers to the general topic or main theme of the film. For example, a "depressing" film that revolved around the topic of suicide would fall into this category. Critics attempting to partition the market for films are likely to refer to these dimensions when evaluating works because they are central aspects of all films. Indeed, compared to dimensions such as direction or characterization, these three dimensions were described the most frequently by critics in their reviews.

The adjectives were then used to calculate the amount of synonym overlap between films in the sample. In the present analyses, we employ a dichotomous measure of synonym overlap—if a pair of adjectives has any common synonym, the pair is assigned a similarity score of 1; otherwise, the score is 0. The similarity score for each pair of films is a continuous variable with values running from 0 to 1. The score measures the proportion of adjective pairs that have similarity scores of 1 out of all possible adjective pairs for a certain dimension between two films. It is computed using the following formula:

$$S_{ij} = \frac{\sum_x \sum_y A_{xy}}{N_i * N_j},$$

where S_{ij} is the similarity between film i and film j , A_{xy} is the similarity between the x th adjective of film i and the y th adjective of film j , N_i is the number of adjectives used by the reviewer to describe film i , and N_j is the number for film j .

3.3. Developing cluster profiles

Before developing the cluster profiles, we first select relevant characteristics by which observations could be grouped. There are a number of possible ways to categorize the critical reviews for comparison. For instance, we could classify each review according to the year in which the reviewed film was distributed and compare the cluster profiles of reviews over time. Or we could group reviews by specific critics and compare the cluster profiles of the different critics in our sample. For the current study, we chose to classify reviews by the genre of the reviewed film, the periodical in which the review appeared, and the time period in which the review was published. Such a classification scheme allows us to examine whether there are salient differences in schemas by genre, by periodical (*New York Times* versus *Variety*), and finally by time period, which we split into early (1946–1961) and late (1964–1982) periods.

For the cluster analyses, we used clustering algorithms available through (SAS Institute Inc., 1989). In the initial step of the cluster analysis, we drew on Ward's clustering algorithm to determine cluster centers. These centers were then used as the initial seeds for PROC FASTCLUS, SAS's version of the k -means algorithm.

We use this clustering procedure to construct cluster profiles for each genre–periodical–period set of reviews. Fig. 8 provides an illustration of the differences between the cluster profiles of different types of distributions. This figure plots the percent variance unexplained at different numbers of clusters, ranging from low to high, for four sets of reviews in our sample. The sets corresponding to the simple and robust distributions by definition explain the most variance at lower numbers of clusters. The robust curve decreases more sharply than the simple curve as the cluster numbers increase, however. Therefore, at higher cluster numbers, the robust set has considerably more explained variance than the simple set. Fig. 8 also shows that, like the robust set, the complex set has high explained variance at high cluster numbers. At low cluster numbers, by comparison, the complex set has considerably less explained variance than the robust set. Not surprisingly, the unpartitioned set explains the least of the four sets for both low and high numbers of clusters.

While this figure provides a way to compare a limited number of schemas, it does not provide a particularly parsimonious way to represent a population of schemas. One way to simultaneously represent a large number of schemas is to depict the schemas in a two-dimensional space like that depicted in Figs. 4 and 6. That is, one plots

Fig. 8. Cluster profiles.

the percent variance explained by a low number of clusters on one axis and the percent variance gained by increasing the number of clusters on the second axis. The problem with this type of representation is that there is a certain arbitrariness in deciding upon the number of clusters that should be considered “low-*m*” on the vertical axis and “high-*m*” on the horizontal axis. One reasonable approach to determining the values for the axes is to identify the optimal cluster number for each distribution within a set and then set the “low-*m*” and “high-*m*” at the extremes of these. There exist a variety of methods for identifying the number of clusters that most appropriately represent the underlying data. These so-called “stopping rules” typically evaluate some measure of the goodness of a cluster solution and identify the number for which this measure is optimized (Gordon, 1999). In a review of more than 30 stopping rules, Milligan and Cooper (1985) found Calinski and Harabasz’s (1974) pseudo *F* statistic and Sarle’s (1983) cubic clustering criterion to be two of the more effective for assessing the number of clusters present in a set of simulated data. Obviously, such statistics by themselves do not capture as much information as a cluster profile and, thus, by themselves cannot be used to distinguish schemas as simple, complex, unpartitioned, or robust. However, these methods can be effectively used to suggest the “low-*m*” and “high-*m*” values on our axes. We applied these statistics to every one of our genre-periodical-period matrices, and across all of the matrices, the largest number of clusters needed to adequately represent the underlying data was six according to both statistics. The lowest suggested number of clusters was two.⁹

Fig. 9 illustrates the location of the critical schemas in the two-dimensional space. This figure enables us to detect underlying trends in the cluster profiles of critical reviews by periodical. For instance, reviews published in *Variety* during the early period tend to concentrate in the unpartitioned quadrant of the graph. Reviews from the early period of the *New York Times*, by comparison, have high explanatory power for their two-point distributions and tend to be either simple or robust. This suggests that, during the early period of our study, critics belonging to the different periodicals envisioned and performed their roles as critics in fundamentally different ways. While *Times* critics created a simple and straightforward schema for distinguishing between different types of films, critics for *Variety* did not create any clear differentiation. As Fig. 9 shows, this difference between reviews in the two periodicals becomes more blurred in the later period.

Some of the content-based changes seem noteworthy. For example, one of the largest shifts is the movement of the *New York Times*’ critical schema for war films. In the early period, war is relatively simple genre; in the later period, it is more complex. This seems appropriate given the changing attitudes toward war after Vietnam. The *Variety* reviews also made a similar shift, moving down and to the right. However, at least given where the line between unpartitioned and complex is currently drawn, *Variety*’s schema in the later period falls on the unpartitioned side of the

⁹ A low value of two implicitly raises the issue that these stopping rules are not equipped to provide guidance regarding whether or not the data should even be partitioned in the first place. Since we need a minimum of two clusters to construct cluster profiles, however, this is not a problem for the analyses at hand.

Fig. 9. Distribution of cluster profiles.

dividing line. There are also is a clear tendency for the “lighter” genres—comedy, musicals, and romance—to be further to the left on the horizontal axis, implying that the critical schemas associated with these genres tend to be either simple or unpartitioned. That is, these genres do not seem to lend themselves to as richly elaborated critical schemas as some of the other genres.¹⁰

4. Conclusion

The purpose of this paper has been to develop a methodology for comparing the critical schemas that serve as the lens through which products are viewed by consumers.

¹⁰ As stated earlier in the discussion of clustering methods, it is often useful to compare the results of a method to those of others when conducting cluster analyses. We therefore, also constructed cluster profiles using the partitioning method PAM on the same data set. We chose this alternative partitioning method because, as stated earlier, the specific goal of constructing cluster profiles would be better met through a partitioning rather than a hierarchical method. PAM minimizes the sum of distances of each object to its cluster’s medoid and is accordingly more robust than *k*-means with respect to outliers. When mapping the critical schemas in the two-dimensional space according to the results found through PAM, we find that 76.5% of the observations remain in the same quadrant they were assigned through *k*-means. This percentage is roughly three times higher than what one would expect due to random assignment of genres to quadrants. One might reasonably observe that random chance does not seem like a particularly stringent baseline against which to evaluate the agreement of two clustering methods. However, in the absence of an obvious alternative baseline, we believe that we can do little more than report the percentage.

One limitation of the methodology seems especially important to note. While this methodology provides a way of formally assessing differences between schemas, it does not offer a statistical basis for comparison. That is, the methodology does not allow one to decide whether the differences between two schemas are greater than what one would expect under chance. At least at the present time, we do not see a way to make such statistical inferences given the lack of any basis for making distributional assumptions about the language that critics employ.

However, even without a statistical basis for discriminating between schemas, we believe this methodology has the potential to illuminate the impact of such schemas on market outcomes. Indeed, there are a number of market outcomes that could be affected by the critical schemas to the extent that they become the lens through which consumers view products. For example, one might hypothesize that the more complex the schema, the more distinct niches will exist for separate products. As a result, a complex schema may enable differentiation that in turn reduces competition among producers in a market.

We have already started some preliminary investigations into the relationship between critical schemas and the artistic and market success of films that occupy positions within those schema. Specifically, with the results depicted in Fig. 9, we use a film's genre to assign the film to one of four schema types—simple, robust, complex, or unpartitioned. A film can actually receive two designations—one from the *Times* critic and one from the *Variety* critic. For example, if a film is a late period war production, then the film would be assigned a complex designation due to the *New York Times* critics' schema for late period war films and an unpartitioned designation due to the *Variety* critics' evaluation. Measures of artistic and commercial success can then be regressed on eight indicator variables for review source (*New York Times* or *Variety*) and schema type (simple, robust, complex, or unpartitioned).

Preliminary analyses show that the categorizations of the *New York Times* critics seem to explain more variance in artistic and commercial success than the categorizations of the *Variety* critics. This is not surprising given that the audiences for the two periodicals differ in terms of the amount of information they require for decisions regarding consumption. A major audience for *Variety* consists of film distributors whose main concern is which films will bring in the highest amount of revenue in terms of ticket sales. One might expect categorizations of films geared towards this audience to be less informative than films geared towards a broad and diverse group of movie-goers who are trying to learn about which films will be the more enjoyable for them. Because the *Times* must speak to an audience with wildly differing preferences and demands, its reviews are more likely to capture important differences among films. The analyses also show that films belonging to simple and complex genres experience more commercial and artistic success than films belonging to unpartitioned and robust schemas. Given that simple and complex schemas are less ambiguous schemas than either unpartitioned and robust schemas, this result is suggestive. It indicates that the clearer a schema—regardless of its level of complexity—the more value that it imbues in the products that are perceived and experienced through it. Obviously, such an inference can only be regarded as tentative at best. The *New York Times* and *Variety* critics are not the only critics providing

lenses for audiences to view films, and our analysis does not include a rigorous set of control variables. However, since the primary purpose of this paper is to develop a methodology for evaluating critical schemas, we believe that a more detailed focus on the analysis would detract from the central objective of the paper—developing a methodology for the comparative analysis of critical schemas.

As we contemplate future development of this methodology, we believe that there are two particularly important extensions to pursue. One of these extensions is methodological; the other is substantive. The methodological issue relates to our particular measure of distance. While we believe that the use of key adjectives is a reasonable starting point for measuring distance, we also believe that many developments in linguistics might usefully be employed for a richer measure of distance. There exist a variety of grammars—for example, dependency grammar and systemic functional grammar (see Jurafsky and Martin (2000))—that provide a basis for isolating key themes in texts. By extracting and dimensionalizing these themes, we believe that it will be possible to obtain an even better representation of the linguistic landscape.

The substantive extension involves a move from comparing schemas within a given market to comparing schemas across markets or exchange domains. Podolny and Hsu (2002) discuss systematic differences in the quality schemas underlying artist worlds, professional arenas, and product markets. Cross-market comparisons would provide a basis for evaluating the extent to which a given exchange domain is becoming more artistic, more professional, or more of a conventional product market. Such analyses would be of particular interest in such areas as film, which seems to constitute an exchange domain at the border of art world and product market. Indeed, there are numerous examples of objects that seem to shift between art and market product. Wine, quilts, and clothing design are just a few of the more obvious examples. Through longitudinal analyses, one might even be able to specify the extent to which an exchange domain is becoming more or less “art-like.” However, regardless of whether these extensions or others are pursued, subsequent research will hopefully document the usefulness of this methodology in uncovering how critical schemas impact on matching processes in the market.

References

- Adams, L., 1990. *The Wines of America*, third ed. McGraw Hill, New York.
- Banfield, J.D., Raftery, A.E., 1993. Model-based Gaussian and non-Gaussian clustering. *Biometrics* 49, 803–821.
- Becker, H., 1990. *Art Worlds*. University of California Press, Berkeley.
- Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3, 1–27.
- Davis, G.F., Diekmann, K.A., Tinsley, C.H., 1994. The decline and fall of the conglomerate firm in the 1980s: the deinstitutionalization of an organizational form. *American Sociological Review* 59 (4), 547–570.
- Everitt, B.S., Landau, S., Leese, M., 2001. *Cluster Analysis*, fourth ed. Oxford University Press Inc, New York.
- Gordon, A.D., 1999. *Classification Monographs on Statistics and Applied Probability*, second ed. 82 Boca Raton, Chapman & Hall.

- Greenfeld, L., 1989. *Different Worlds: A Sociological Study of Taste, Choice, and Success in Art*. Cambridge University Press, Cambridge, England.
- Hirsch, P.M., 1986. From Ambushes to Golden Parachutes: corporate takeovers as an instance of cultural framing and institutional integration. *American Journal of Sociology* 91 (4), 800–837.
- Jurafsky, D., Martin, J.H., 2000. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall, New York.
- Kaufman, L., Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Leifer, E.M., 1988. Trails of involvement: evidence for local games. *Sociological Forum* 3 (4), 499–524.
- MacQueen, J., 1967. Some Methods For Classification and Analysis of Multivariate Observations. In: *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* 1, pp. 281–297.
- McQuitty, L.L., 1957. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. *Educational and Psychological Measurement* 17, 207–229.
- Milligan, G.W., 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 45, 325–342.
- Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* 50, 159–179.
- Padgett, J.F., Ansell, C.K., 1993. Robust Action and the rise of the medici, 1400–1434. *American Journal of Sociology* 98 (6), 1259–1319.
- Podolny, J.M., 1993. A status-based model of market competition. *American Journal of Sociology* 98 (4), 829–872.
- Podolny, J.M., Hsu, G., 2002. Quality, social exchange, and knightian uncertainty. In: Raub, W., Buskens, V., Snijders, C. (Eds.), *Review of Organizational Sociology*. JAI Press, Greenwich, CT.
- SAS Institute Inc., 1989. *SAS/STAT User's Guide, Version 6*. Cary, NC, SAS Institute Inc, 823–850.
- Sarle, W.S., 1983. *Cubic Clustering Criterion*, SAS Technical Report A-108 Cary, NC, SAS Institute Inc.
- Scott, A.J., Symons, M.J., 1971. Clustering methods based on likelihood ratio criteria. *Biometrics* 27, 387–398.
- Sneath, P.H.A., 1957. The application of computers to taxonomy. *Journal of General Microbiology* 17, 201–226.
- Sokal, R.R., Michener, C.D., 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38, 1409–1438.
- Tuma, N.B., Hannan, M.T., 1984. *Social Dynamics: Models and Methods*. Academic Press, Orlando, FL.
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58, 236–244.
- White, H.C., 1981. Where do markets come from. *American Journal of Sociology* 82, 517–547.
- White, H.C., 1993. *Careers and Creativity: Social Forces in the Arts*. Westview Press, Boulder, CO.
- White, H.C., White, C.A., 1992. *Canvases and Careers: Institutional Change in the French Painting World: With a New Forward and a New Afterward*. University of Chicago Press, Chicago.
- Zuckerman, E.W., 1999. The categorical imperative: securities analysts and the illegitimacy discount. *American Journal of Sociology* 104 (5), 1398–1438.
- Zuckerman, E.W., 2000. Focusing the corporate product: Securities analysts and de-diversification. *Administrative Science Quarterly* 45 (3), 591–619.