Research paper

Developing a digital marketing tool for ethnic ventures' mixed business model and market-shaping: A design scientific approach of web demographics

Taewon Suh [a],[*], T. Edwin Chow [b]

[a] *McCoy College of Business, Texas State University, USA*
[b] *Department of Geography, Texas State University, USA*

ARTICLE INFO

ABSTRACT

Based on the advanced market conceptualizations and the emergence of market-shaping strategy and mixed business models, this study develops a theoretical definition and a practical solution of digital marketing for ethnic ventures' mixed business model and market-shaping strategy. This study argues that building an ethnic foothold can feasibly empower ethnic ventures' market-shaping over mass customization and customerization and thus allows the ventures to grow fast through their mixed business model combining business-to-business and business-to-consumer. Subsequently, for the main part, adopting a design scientific approach and relying on web demographics' interoperability and geographic reference, we design and test a procedure of market-shaping strategy for ethnic ventures. Exemplified by a data subset collected from people-finder websites regarding Vietnamese Americans in Texas, our process of mixed business model and market-shaping strategy uniquely involves three stages: (1) *acquisition and identification*, (2) *description and secondary clustering*, and (3) *conscription and network-targeting*.

1. Introduction

The notion that the traditional way of target marketing has lost its relevance may be due to the over-emphasis on the homogeneity of customer segments and the contradictorily increased diversity and dynamics of customer needs and behaviors (Firat & Shultz, 1997; Harrison & Kjellberg, 2010; Yankelovich & Meer, 2006). We assert that this notion can be extended to the idea of *mixed business model* for new ventures and fast-growing companies. It is excessively simplistic to dichotomize markets into business-to-consumer (B2C) and business-to-business (B2B) (Leek & Christodoulides, 2011). Many fast-growing companies develop a mixed business model to maximize both profitability and scalability. Recognizing the role of digital marketing in multiple business models across B2B, B2C, and business-to-business-to-consumer (B2B2C), this paper particularly gives its focus on the design of digital marketing solution for the companies with mixed business models and their market-shaping. *A B2B2C is a mixed type of business model in which the company indirectly assesses the consumer market* via *another business.* Over time, targeting marketing activities at the end consumer and intermediary simultaneously, the company offering the

service may gain direct access to consumers (Iankova, Davies, Archer-Brown, Marder, & Yau, 2019).

While the notion of mixed business model can be thought as an invention of the market, it has been documented that the homogeneity of groupings in traditional target marketing strategy is not a naturally occurring market phenomenon (Quinn, Hines, & Bennison, 2007; Yankelovich & Meer, 2006). New ventures and fast-growing companies may need more than a managerial assessment to identify a market foothold and deploy their products into a more complex market structure with the mixed business model. In order to expound upon mixed business model and the appropriate marketing tool for implementation, this study particularly connects the issue to a recent development in market conceptualizations, known as *market-shaping*. The research of market-shaping progressively recognizes markets as socially constructed (Kjellberg & Helgesson, 2006; Harrison & Kjellberg, 2010; Nenonen, Storbacka, & Windahl, 2019). Networks, systems, or ecosystems are used to replace markets in the literature (Iansiti & Levien, 2004; Johanson & Vahlne, 2011; Mattsson & Johanson, 2006). New ventures and fast-growing companies do not shy away from taking a mixed business model and shaping their market. Some businesses are able to

---

uniquely define their market and sell products to both businesses and individual consumers, just as Amazon has both business customers and its own retail (Iankova et al., 2019; Ries, 2011).

Since markets can be understood as a context for value co-creation, companies are oriented to understand how players in a larger system can co-create value and recognize the institutional arrangements that guide each behavior in the process (Nenonen et al., 2019, p. 619). A few new target marketing approaches have emerged to absorb the market-shaping orientation as a logical step towards gaining a greater understanding of particular markets (Harrison & Kjellberg, 2010; Quinn et al., 2007; Venter, Wright, & Dibb, 2015). Following these precedents, we aim to develop a new procedure of mixed business model for market-shaping using big data analytics. We define such procedure as *market-shaping* target marketing wherein the firm can actively learn more about the given market structure and facilitate a certain pathway of incremental targeting to market penetration in the mixed business model. As an interactive and emergent process, market-shaping target marketing works to develop an "actionable map of the market" (Harrison & Kjellberg, 2010). We distinguish the market-shaping approach from the traditional approach, *market-driven* target marketing (Jaworski, Kohli, & Sahay, 2000), which is based on clustering the existing market and describing targeted segments with the managerial objective to identify homogenous groups within a defined market.

Successful market-shaping brings forth market innovation, which involves the introduction of new devices and the emergence of new solutions (Kjellberg, Azimont, & Reid, 2015; Vargo, Wieland, & Akaka, 2015). New ventures and fast-growing companies commonly endeavor to maximize their scalability, often blurring the demarcation between B2B and B2C (Ries, 2011). We believe that this recognition may expedite more research in the prescriptive mode of design science than in the descriptive mode of normal science (Simon, 1996). While normal science, driven by a quest for truth, is interested in "what is" which leads to descriptive knowledge, artificial or design science is about a quest for utility questioning "what can be" leading to prescriptive knowledge. It has been discussed that design science research can potentially fill up the theory-practice gap (Benbasat & Zmud, 1999; Hevner, March, Park, & Ram, 2004; Rosemann & Vessey, 2008). However, in the research field of marketing, we find a scarcity of design scientific approach through which the prescriptive mode of research can be carried out to develop marketing devices and solutions. To our surprise, there are only a few exceptions in the extant literature (Beloglazov, Benergee, Hartman, & Buyya, 2014; Hevner et al., 2004), which are not so widely recognized. The marketing field has not reached an understanding about what defines a clear knowledge contribution from a design scientific approach. This shortcoming has limited the adoption and acceptance of design science.

Against this backdrop, we propose a design scientific approach for the development of digital marketing solution, which is a market-shaping tool to mitigate the current issues and problems of new ventures and fast-growing companies in the context of mixed business model. It is acknowledged that design science research is also theory-generating by integrating methods of grounded theory development with established design science methodology (Beck, Weber, & Gregory, 2013). When it comes to the topic of mixed business model and market-shaping, the design scientific approach would be more relevant for both academia and practice to increase the value and usefulness of research.

1.1. Mixed business model and ethnic ventures' market-shaping

As the combination of B2B and B2C, a B2B2C business model necessitates the first B to develop relationship with its intermediary businesses (the second B's) and their consumer customers (C's). The relationships of B2B2C business model heavily rely on the firm's network which is a complex and dynamic network of relationships based on information and computer technology (Gulati, Nohria, & Zaheer, 2000). Therefore, the role of digital marketing is critical in the mixed

business model (Iankova et al., 2019; Lacoste, 2016). We attempt to relate the development of B2B2C business model to market-shaping in which firms need to understand how players in a larger system can co-create value. The context of ethnic ventures was appropriately chosen for this study to present the development of the proposed market-shaping solution. With concentrated locations of immigrant groups where a variety of enterprises serve and source from their ethnic market, ethnic ventures take several advantages and attempt to recognize opportunities in their enclave economies (Portes & Zhou, 1992. Bird & Wennberg, 2016).

With enhanced cross-cultural experience, migrants and expatriates can better identify profitable opportunities and exhibit a higher level of entrepreneurial activity (Masurel, Nijkamp, & Vindigni, 2004; Beckers & Blumberg, 2013; Ojo, Nwankwo, & Gbadamosi, 2013; Ram, Jones, & Villares-Varela, 2017). Ethnic ventures often choose a strategy that focuses on their ethnic enclave rather than a dominant market strategy (Beckers & Blumberg, 2013; Bird & Wennberg, 2016; Neumeyer, Santos, Caetano, & Kalbfleisch, 2018). This niche market strategy may be attributed to both the advantages the ventures have in their ethnic communities and the disadvantages they encounter outside of their ethnic communities. Ethnic ventures take several advantages in their enclave economies with concentrated locations of immigrant groups where a variety of enterprises serve and source from their own ethnic market (Portes, 1981; Bird & Wennberg, 2016). The literature sorted out a list of the advantages, including access to unorthodox means of financial capital (Portes & Zhou, 1992), community sponsorship (Chaganti, Watts, Chaganti, & Zimmerman-Treichel, 2008; Greene, 1997), access to co-ethnic networks (Butler & Greene, 1997), and the potentiality for a cheaper alternative to vertical integration (Wilson & Portes, 1980). The ethnic enclave arguments indicate that immigrant entrepreneurs fundamentally consider "the degree of focus on the ethnic market and the extent of ethnic resources used" (Ndofor & Priem, 2011, p.793; also, Bird & Wennberg, 2016).

Targeting ethnic groups is advocated because they are marketing relevant (Pires, Stanton, & Stanton, 2011; Rossiter & Chan, 1998). The literature supports that the aspects of human capital can determine the extent to which an entrepreneur recognizes opportunities (Gedajlovic, Honig, Moore, Payne, & Wright, 2013; Venkataraman, 1997). The cognitive behaviors of entrepreneurs are bounded by their ability to gather appropriate information and the knowledge of how to process information (Woo, Folta, & Cooper, 1992). The above arguments overall reveal the importance of ethnic market strategy.

1.2. Our approach: Web demographics and design science research

Utilizing personal microdata available on the web, or web demographics, one can gather basic attributes of individuals distinguishable to an ethnic population down to the neighborhood level (Chow, 2013; Chow, Lin, & Chan, 2011). Our approach in web-demographic, target marketing strategy corresponds to the marketing-shaping aspect of the new market conceptualizations. Facilitating web demographics, this study develops a design scientific procedure with which ethnic ventures can find profitable opportunities. By scaling, measuring, and documenting information from individual consumers in a society, web demographics (i.e., large quantities of demographic micro-data available on the Internet) provide the fundamental unit to integrate other channels of data [0]. The web demographics at the individual level are flexible in spatial scales to be synthesized with other big data at the individual level, such as social media and travel trajectory enabled by Global Positioning Systems (GPS), or conventional geodemographic data, such as block groups or census tracts, at the aggregated level. In light of big data empowerment, web demographics enable data-driven marketing analytics and spawn innovative applications.

The following sections introduce a three-stage approach based on the widely-accepted guidelines and sequences of activities in the design science research (Beloglazov et al., 2014; Hevner et al., 2004;

Offermann, Levina, Schönherr, & Bub, 2009; Peffers, Tuunance, Rothenberger, & Chatterjee, 2008): (1) *Problem Definition*, (2) *Design and Development*, and (3) *Evaluation*. In problem definition, we identify and define the problems in ethnic ventures' target marketing strategy for their mixed business model and justify the value of the proposed solution. In the stage of design and development, we craft the course of strategy by determining its desired functionalities and architecture. Web demographics is presented as an alternative data source of market segmentation. Accordingly, we configure the three components of target marketing strategy using web demographics. In the final stage of evaluation, we measure how well the designed procedure supports the solution to the problem and demonstrates the use of the design scheme. We accomplish the evaluation by arranging an expert survey and by means of a coverage and accuracy test of web demographics.

This study's main contribution concerns the three stages of market-shaping in the solution process for ethnic ventures: (1) *acquisition and identification*, (2) *description and secondary clustering*, and (3) *conscription and network-targeting*. This three-stage process of market-shaping based on web demographics illustrates the incremental progress of the target marketing process. Drawing on the experience gained from their relationship with initial foothold, which is a conscripted segment, ethnic ventures are able to map their market by adding more sub-segments, aggregating more customers to the segment or attempting an overspill to other segments. Another trait of market-shaping segmentation concerns the strategic activities used to adjust the market to the envisioned map by stabilizing the relationship between the foothold and the whole market, which is denoted by mapping (Harrison & Kjellberg, 2010).

2. Problem definition

2.1. Need for ethnicity-based target marketing

Concerning the practical matters in using web demographics for ethnic ventures, we mainly give our focus on a contemporary approach in the combination of digital marketing and market segmentation. An effective market segmentation, which is *dynamic, appropriate*, and *substantial*, should focus on a simple set of issues and be continuously re-explicated to maintain its relevance and strategic momentum (Brotspies & Weinstein, 2019; Pires et al., 2011; Yankelovich & Meer, 2006). This notion corresponds to several examinations to enhance the practical value of target marketing (Lee, Fairhurst, & Dillard, 2001; Hassan, Craft, & Kortam, 2003; Yankelovich & Meer, 2006; Pires et al., 2011). Also, the concept of ethnicity-based marketing is central to our design scientific approach for ethnic ventures' market-shaping. Against the backdrop, therefore, we define ethnicity-based target marketing as *a strategic method in target marketing process that uses minority ethnic group as a foothold for growth*. Its primary justification may rely on the advent of big data analytics. Ethnic groups refers to "broad groupings of individuals on the basis of both race and culture of origin" (Phinney, 1996, p. 919). The view that a unique ethnic group of customers can be distinguished from other groups of different ethnicities has been suggested by cultural deterministic views (e.g., Triandis, 1996). Identification of minority ethnic groups is relevant to business practices because their behavior is predictable based on group mechanisms and synergies.

We deem that observable ethnicity may pose an interesting approach to overcome the over-simplification of the singular use of descriptive variables in market segmentation. The usefulness of using observable ethnicity as a segmentation basis stems from the evolving nature of international customers' ethnic identity (Pires et al., 2011). While culture guides customers' behavioral patterns (Aaker & Maheswaran, 1997; Penaloza, 1994), a cultural compound made through ethnic identity with other varied cultural aspects may have an influence on the multiple layers of minority customers' cognition, emotion, and behavior. Customers' ethnic identities are evolving into an amalgam of multiple cultural factors through an ongoing process (Lee et al., 2002).

If ethnic groups are narrowly defined and become invisible when lumped into other classifications in official statistics, firms should cautiously use ethnicity-based target marketing strategy. Invisible groups tend to accommodate inferences of excessive smallness or lack of importance that can combine with managerial preconceptions that ethnicity-based target marketing strategy is too costly or prone to social retaliation (Cui & Choudhury, 2002). In our attempt to overcome the deficits of the traditional target marketing, we will recommend a process of ethnicity-based target marketing strategy using web demographics that can provide multiple solutions to the newer challenges in the target marketing process within the globalized markets.

2.2. New technique for market-shaping

In the Geographic Information Systems (GIS) field, web demographics are useful in analyzing the spatio-temporal distribution among minorities and ethnic groups (Chow, 2013). This conceptual and technological development also supports the emergence of ethnic market segments, which can be attributed to a macro change in population composition. Since 2011, 50.4% of the U.S. newborn babies, who were one year old or younger, were from ethnic minorities. This trend narrows the gap between the notions of "majority" and "minorities" and is predicted to redefine the definitions of those terms within a few decades (U.S. Census Bureau, 2012). Ethnic minority populations' high growth rate in the U.S. alone justifies ethnicity-based segmentation (Wyner, 2009). The literature also argues that the US ethnic minorities fit most of the key criteria of segmentation (Pires et al., 2011; Rossiter & Chan, 1998).

The primary data source of traditional target marketing for geodemographic analysis is census data, which is commonly aggregated to a broad geographic area such as census block or block group. Depending on whether the decennial census or rolling annual American Community Survey (ACS) desirable for the ethnic market segmentation, some socioeconomic and demographic attributes are only available at certain spatial scale to protect personal identity. For example, median household income is not available at census block level. While race is generally available at fine spatial scale, ethnicity data (e.g., Japanese, Cuban) is often available only at a very coarse scale (e.g., county), which makes strategic marketing at the neighborhood level impossible. Moreover, decennial census and ACS are conducted in a wide time-span from one, three, five to ten years with varying degree of sampling bias and temporal currency. Thus, the utility of census data is limited in target marketing process.

Concerning our attempt to proactively use web demographics for ethnic ventures' market-shaping, we need to define significant issues in the current market segmentation practices. Examining the literature of market segmentation, we precisely identified three categories of anticipated and envisaged changes in the practice: namely, *dynamic segmentation, appropriate segmentation*, and *substantial segmentation*. These three categories reflect the dissatisfactions present with current practices and hold the promise for providing marketers with the information and knowledge needed to make better decisions in the segmentation process.

First, a new practice of market segmentation requires *dynamic segmentation*. While market segmentation is indispensable in identifying homogenous segments, firms may seek to grow their businesses beyond homogenous groups (Raynor & Weinberg, 2004). In a dynamic process, they can first launch a strategic solution in one segment and then build up the market to compete for another, bigger segment from the position of the built advantage of the first segment. Therefore, it is suggested that market segmentation should become more adaptable and less costly to accordingly respond to market changes. For example, marketers can direct its course simply on one or two issues, which are redrawn when the issues have lost their relevance (Yankelovich & Meer, 2006).

Second, market segmentation requires *appropriateness* from the perspective of management. Market segmentation may do itself harm by failing to make itself understood by its senior management. According to Yankelovich and Meer (2006), there are three types of failure in the

practice of segmentation: (1) excessive interest in consumers' identities, (2) little emphasis on consumers' actual behaviors, and (3) excessive absorption in the technical details. These potentially lower the levels of appropriateness and ease of implementation of the resulted segments. We can also find an apparent gap regarding segmentation between actual business practices and academic research (Natter, Mild, Wagner, & Taudes, 2008). In practice, firms may tend to consider appropriateness and ease of implementation to make market segmentation decisions (Wind & Cardoza, 1974). If the segments seem inconsistent with the work experience of managers who cannot grasp what the segment means, how they were derived, and how related programs can be implemented, the research yielded is unlikely to be accepted and applied (Wind & Cardoza, 1974). Thus, segmentation needs to be intuitively understood by internal constituencies concerning varied aspects of marketing. Failure to do this may cause firms to improperly focus their resources incorrectly on only one kind of strategy. This line of argument often leads firms to a simpler, structured process of segmentation.

Third, effective segmentation may generate substantive homogenous segments (Dibb & Simkin, 2009). *Substantiality* focuses on a target market, which is large enough to be a potential profit generator (Wind, 1978). In the globalized marketplace, the enhanced heterogeneity of customers prompts a firm to seriously consider market segmentation and its substantiality (Storbacka, 1997). Pires et al. (2011) argue that the substantiality criterion is an assessment of value relative to the business objectives prior to segmentation decision. Often, its treatment in the literature is inadequate, still being an ex-post consideration and dependent on forecasting quality.

2.3. Summary: Research problem

We previously argued that the emergence of big data analytics could lead to a solution to the problems and issues in ethnic ventures' B2B2C business model and marketing-shaping. We redefine the circumstantial problem into the unique research problem to reveal the value of our solution. First, a business model innovation of ethnic ventures requires dynamic market-shaping. While target marketing strategy is indispensable in identifying homogenous segments, marketers need to understand that their firms seek to grow their businesses, perhaps beyond the homogenous groups. Second, their business model innovation requires appropriateness from the perspective of management. Marketing strategy may do itself harm by failing to make itself understood by management. Third, effective schemes in a new business model may generate a substantive customer base. Therefore, based on the requirements, we propose the use of web demographics for ethnic ventures' mixed business model to mitigate the generic issues and problems of ethnic ventures. Mixed business model can provide economic opportunities created for the business, which helps access the consumer market. We believe that ethnicity-based market-shaping can potentially overcome the pitfall of a B2B2C business model, which is the lack of push and distribution from its business partners.

3. Design and development

3.1. Web demographics

Using people-finder websites as data gateways, Chow et al. (2011) proposed a framework to conduct a "web census" for population groups with highly predictive surnames based on surname analysis. In 2009, a pilot study used a list of 91 surnames highly predictive to Vietnamese Americans (VA). In Texas, it returned about 81,354 and 20,957 individual records from the Intelius (www.intelius.com) and WhitePages (www.whitepages.com) websites, respectively (Chow, Lin, Huynh, & Davis, 2012). After removing the erroneous and fictitious records, the combined samples represented about 51.2% of the 190,647 VA (Census Bureau 2013). The resulting web demographics can be analyzed in a Geographic Information System (GIS) and present an innovative

enumeration strategy that opens up new frontiers of research in marketing science. While the precise criteria and definition of big data is evolving, important aspects of big data include its volume, variety, velocity, and value (Chang, Kauffman, & Kwon, 2013; McAfee & Brynjolsfsson, 2012). It is perhaps useful to examine the characteristics of web demographics in terms of these four aspects.

In theory, web demographics are proportional to the human population. Thus, the number of individuals solicited by web demographics has a finite volume. There is not an official count of web demographics, since bits and pieces of personal microdata are scattered across a variety of public and private sources. People in developed countries are more likely to leave behind a digital trace than those in developing countries. As the infrastructure of Information Communication Technologies (ICT) continues to improve in developing countries, it is expected that the total count of web demographics will approximate the world population. While there is a finite limit in terms of human population, there are infinite geographic, demographic, social, economic, and behavioral attributes that can be associated with each person. Web demographics are highly scalable because individuals are the fundamental unit of data analytics in social and marketing science. This implies that web demographics can function as the primary key in the large human data universe and can grow organically by joining with sources of big data that can be associated with human activities (e.g., tweets, calls, emails, web navigation history, keyword search, etc.). Thus, the volume of web demographics, in addition to the steady increase of the human population, is largely dependent on the potential attributes that can be recorded and traced to an individual. It is worth noting that many of these attributes are spatio-temporal dynamic so that the volume can grow exponentially if these attributes are archived (e.g., household members at a specific time and place).

Originating from multiple sources, web demographics have a chaotic data structure, and its content contains instant and cross-referenced information (Chow, 2013). A valid record of web demographics may include a name. Names are normally composed of first and family names that are spelled out in full or initialized. On the other hand, other demographic data, such as address and phone number, do not have a clear definition or formal structure. Addresses can range from a full address at the street level to a simple location like city or country. Mentioning city location rather than address is not uncommon in a lot of self-reported user profiles of social media platforms (e.g., Facebook). Some people-finder sites even include relatives or associated members in the household. The content of other big data which is often combined with web demographics is highly varied. It includes pictures, videos, text, and other mixed media information that can be qualitative or quantitative.

Among common sources of big data, the importance of web demographics is two-fold: (1) interoperability and (2) geographic reference. As mentioned, its interoperability can be attributed to the fact that web demographics can be joined with other big data as long as there is a link between a record in the big data and a person. Moreover, web demographics are spatial with a street-level address that can be precisely mapped. These addresses can be address-matched as a point layer in GIS and overlaid with other geographic layers to explore their spatial relationships. Thus, its spatial aggregation is flexible at any scale and can be used with other GIS data for map overlay and spatial analysis. For example, it is possible to explore the spatial association between any big data combined with web demographics and the demographic and socioeconomic attributes inferred from Census data. Thus, web demographics are valuable due to their high scalability and interoperability with other spatial and aspatial data. Web-demographic, ethnicity-based target marketing strategy must overcome some of the challenging issues in the practice of market segmentation through its market-shaping process. We attempt to configure a process of market-shaping, target marketing strategy for ethnic ventures into three stages: (1) *acquisition and identification,* (2) *description and secondary clustering,* and (3) *conscription and network-targeting.*

3.2. Acquisition and identification

The acquisition of web demographics is a five-step process that consists of query, navigation, extraction, post-processing, and mapping (Chow et al., 2011). Most people-finder sites require at least a surname and a reference of location (e.g., zip code; city name) to construct a valid query string. We draw the surname or locational reference from a database through an automated process.

For ethnicity-based target marketing strategy, it is crucial to construct a list of surnames that are highly-predictive of a particular ethnic group (Lauderdale & Kestenbaum, 2000). Based on Zipf's law, the frequency f of surnames is inversely proportional to their ranked order (r) as follow:

$$f(r) \sim \alpha r^{-z} \tag{1}$$

where α is an empirical constant, and z is the Zipf's exponent. In light of Mandelbrot's (1966) generalization, eq. 1 can be revised as:

$$f(r) \sim \alpha (1 + r\beta)^{-z} \tag{2}$$

where β is another empirical constant to fit a specific population group. Word et al. (2011) reported that the cumulative proportion of the population by surname approximates a power law distribution— that about 70 million people (~26% of the U.S. population) share 275 surnames (i.e., 0.00004% out of ~6 million) whereas about 4 million people (~1.5%) take up 64.7% of 6 million surnames. By exploring the empirical relationship between surnames and their population proportion (e.g., eq. 2), a list of "predictive" surnames can be used to enumerate a large sample of an ethnic population for market segmentation.

If multiple people-finder sites were employed to query the web demographics, a record linkage procedure, also known as entity resolution in computer science, would be required to identify duplicated records of the same individual or resolve different attributes of the same individual into a single record. Very often, these duplicated records are the result of repeated enumeration from multiple sites, same name but different addresses (i.e., migrants), or typographic errors in data entry. Using the *Levenshtein* distance to quantify the difference between two names, the Cost Sensitive–Alternative Decision Tree (CS-ADT) was implemented to derive the linkage rule to identify unique or duplicate records (Chen, Borthwick, & Carvalho, 2011). Fig. 1 illustrates the algorithm of CS-ADT employed for entity resolution. After removing the duplicate, erroneous, and fictitious records, individual street addresses are address-matched in a GIS as a point layer.

3.3. Description and secondary clustering

After acquiring and identifying a minority ethnic segment, a further descriptive analysis must be conducted, interpreted, and linked according to expected marketing research and managerial activities. With

Input:

1: database $S = \{(x_1, y_1, c_1), \ldots, (x_n, y_n, c_n),\}$, where $x_i \in X$, $y_i \in \{-1, +1\}$, and $c(i) = c_+ \cdot I(y_i = +1) + c_- \cdot I(y_i = -1)$; c_+ and c_- are misclassification cost associated with positive and negative class examples

2: weights $W = \{W_1^0, \ldots, W_n^0\}$ where $W_i^0 = 1$.   /* uniform distribution initially */

3: $D \leftarrow \{all\ possible\ conditions\}$.

4: $N \leftarrow$ number of iterations

5: $\eta \leftarrow$ smooth factor

Procedure:

6: $r_0 \leftarrow T$ with score $\frac{1}{2} \ln \frac{c_+ \cdot W_+(T) + \eta}{c_- \cdot W_-(T) + \eta}$ /* T: precondition is true for all examples */

7: $P_0 \leftarrow r_0$ /* precondition set */

8: **for** $t: 1 \to N$ **do**

9:    $d_1, d_2 = \text{argmin}(Z)$
        s.t. $Z = \left[2\left(\sqrt{W_+(d_1 \cap d_2) W_-(d_1 \cap d_2)} + \sqrt{W_+(d_1 \cap \neg d_2) W_-(d_1 \cap \neg d_2)} \right) + W(\neg d_1) \right]$
        where $d_1 \in P_t$, $d_2 \in D$

10:    $\alpha_1 = \frac{1}{2} \ln \frac{c_+ \cdot W_+(d_1 \cap d_2) + \eta}{c_- \cdot W_-(d_1 \cap d_2) + \eta} \to d_1 \cap d_2$ /* α_1 is the score associated with $d_1 \cap d_2$ */

11:    $\alpha_2 = \frac{1}{2} \ln \frac{c_+ \cdot W_+(d_1 \cap \neg d_2) + \eta}{c_- \cdot W_-(d_1 \cap \neg d_2) + \eta} \to d_1 \cap \neg d_2$ /* α_2 is the score associated with $d_1 \cap \neg d_2$ */

12:    $r_t \leftarrow d_1 \cap d_2 \oplus d_1 \cap \neg d_2$
        /* $r_t(x)$ is a new splitter node with two associated prediction nodes */

13:    $P_{t+1} \leftarrow \{P_t, r_t\}$

14:    $W_i^{t+1} \leftarrow c(i) \cdot W_i^t \cdot e^{-r_t(x_i) y_i}$ /* update example weights */

15: **end for**

16: **return** Classifier for unseen unlabeled instances:
        $H(x) = sgn(\sum_{t=0}^{N} r_t(x))$

Fig. 1. The CS-ADT algorithm (Chen et al., 2011).

the segment identified, the marketer ascertains whether certain patterns in other characteristics are collected and whether the marketer can re-cluster the ethnic segment into multiple sub-segments based on one or two issues according to a specific pattern meaningful to the marketing program. With the sub-segment formed, the marketer may see how other different characteristics are associated with each sub-segment. In the following section, we demonstrate one example of pattern recognition with the Vietnamese American (VA) segment in Texas.

Ideally, all records of web demographics would have a full name and address at street-level. Some records may contain additional auxiliary information, including age, date of birth, phone number, or names of household members. For instance, age is an important attribute for marketing purposes as it is often desirable to identify the age of an individual or the age structure of a given area to conduct customer profiling. With the VA database brought into the GIS, it is easy to identify the age of an individual observation (i.e., a row in the attribute table) or derive the age structure. In the latter case, discrete individual records in the point layer would be aggregated to a geographic unit first. For example, a census block can be used to conduct a spatial connection with the point layer so that each census block can summarize the number of counts for a specific age group (e.g., 24–29 years old). This process can then be repeated for each target age group in creating a corresponding summary field.

Fig. 2 displays the age structure of all VA in the census block (#18043000) in Austin, Texas. The selected block is located in north Austin (Fig. 2a and b), with several VA-owned international grocery stores and a number of Vietnamese restaurants nearby. The age structure reveals that the three highest frequencies in the age groups are 65 and above (27%), 30–34 (18%), and 35–39 (16%) (Fig. 1d). In an ethnoburb (i.e., a suburban area with a high concentration of ethnic minorities), the bi-modal distribution is a common age structure that reflects the high

concentrations of economically active populations and elderly populations, both of which can benefit from easy access to work or grocery stores with minimal driving. Moreover, Google Earth imagery further reveals a high concentration of the VA population clustered in the apartment complexes within the selected block (Fig. 2c).

Broadly grouped, behavioral perspectives in market segmentation, including a needs-based approach, form segments by looking at behavioral considerations, such as benefits sought, consumption outcomes, use occasions, user status, usage rate, brand loyalty, and switching patterns, and so on (Kotler & Keller, 2007; Lilien, Rangaswamy, & De Bruyn, 2007). Some marketers establish market segments by considering more conventional, descriptive characteristics: geographic, demographic, and psychographic. In the behavioral perspectives, marketers may also decide whether the different descriptive characteristics are associated with each segment. Our proposed ethnicity-based target marketing strategy based on a web census renders a unique method from a behavioral perspective. It utilizes demographics or descriptors in the first two steps and behavioral or other variable categories in the later step, which provides more degree of freedom to marketers.

3.4. Conscription and network-targeting

The dataset from web census can be combined with Volunteered Geographic Information (VGI) data. The prevalence of dynamic services through web and mobile devices nurtures a culture of public participation in producing valuable content. These elements enable a growing movement of active users who respond to a specific crowdsourcing project and contribute relevant geographic observations, known as VGI. For example, OpenStreetMap (OSM) represents one of the earliest attempts to provide a free and editable world map through VGI contributed by millions of users world-wide. We define this process of recruiting
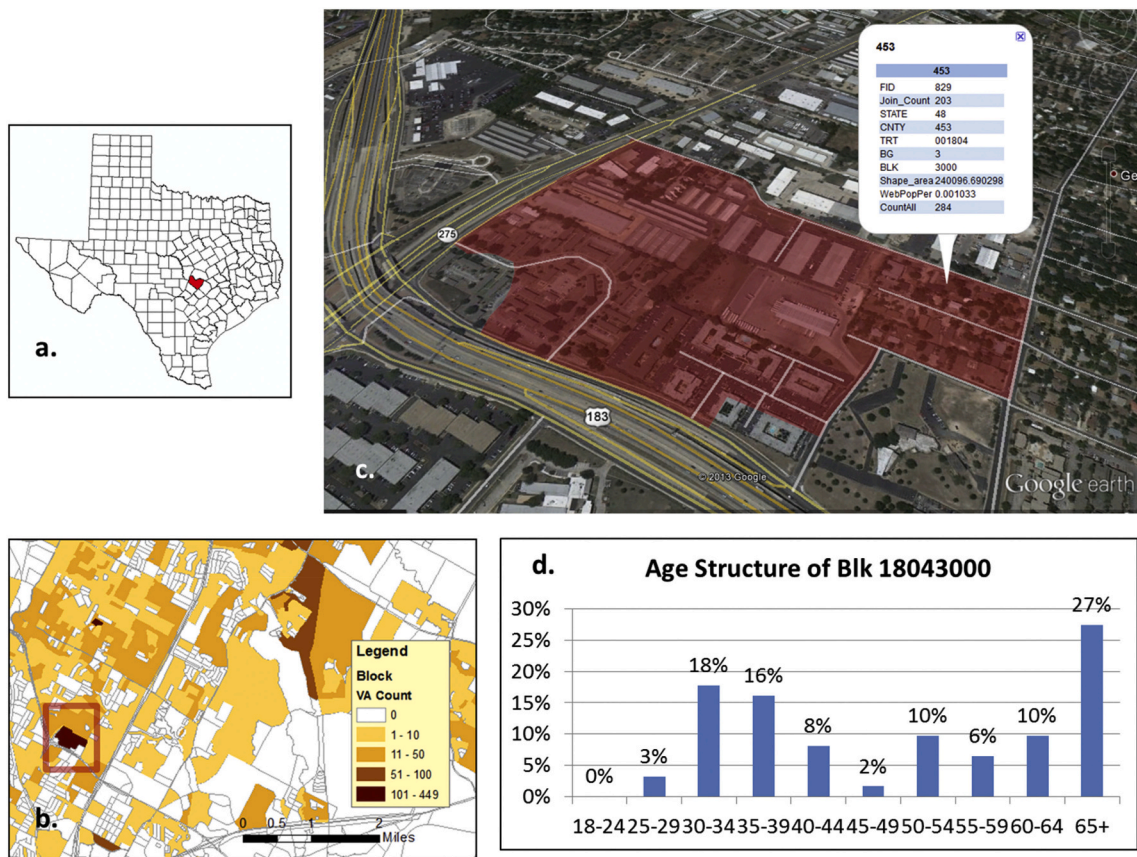

Fig. 2. The age structure of VA in the census block (#18043000) in Austin, Texas. (a) inset map of Austin in Travis county, Texas; (b) the Google Earth imagery of the block; (c) the population count of VA in the neighborhood; and (d) age structure of VA in the selected block.

or selecting more engaging consumers based on VGI as conscription. A conscripted sub-group or sub-segment may be considered as a foothold because the consumers in the group would be more active in responding to marketing programs, as well as in social media.

VGI can be viewed as a user-generated dataset with an explicit geographic reference that can be mapped. VGI is often a subset of the big data because not all data would have an explicit geographic reference. The geographic reference can be the location recorded on the user profile of a social media account, an Internet Protocol (IP) address of the user's computer, or even the x- and y-coordinates when using Global Positioning Systems (GPS) on cellular-based or WiFi-enabled mobile devices. For example, more than one-third of all Twitter tweets can be mapped with a varying degree of spatial precision from street-level to city location (Leetaru, Wang, Cao, Padmanabhan, & Shook, 2013). As the VGI occupies a modest percentage of the 400 million tweets per day, researchers have attempted to explore the use of geotagged tweets to predict box office revenue (Jeon & McSharry, 2013), stock market (Bollen, Mao, & Zeng, 2011), influenza epidemics (Culotta, 2010), and earthquakes (Sakaki, Okazaki, & Matsuo, 2010). Besides tweets, a wide spectrum of VGI has been harvested, such as check-in, gas price, traffic condition, picture, video, bird sighting, damage report, etc. The diverse nature of these VGI presents a rich dataset to be explored.

Since VGI is a user-generated dataset with an explicit geographic reference that can be mapped, the accessibility of segmentation using VGI is still superior to other traditional segmentation techniques. This property of dataset can lower much of the marketing cost when marketers implement their marketing programs based on their segment marketing decisions, which result from segmentation. Although not all data would have an explicit geographic reference, marketers can still find information useful to their integrated marketing program by utilizing multiple touchpoints and media channels.

We can consider the group of entities matched in the process as a sub-cluster that is more active in social media. We define this as conscription. The socially-active sub-segment(s) population P can be a part of market partitioning in the ethnic segment. Using a set of k keywords $W = \{W_1, W_2, \ldots W_k\}$ to query relevant tweets T_W, the proportion of tweets Q that match at least one keyword from all tweets T, so that $Q = (W,T) = \frac{|T_W|}{T}$. For marketing segmentation, one can use regression analysis to estimate the probability of a specific population subgroup i based on the log-odds of P_i and Q as follow:

$$Logit(P) = \beta_1 logit(Q_1) + \beta_2 logit(Q_2) + \ldots + \beta_k logit(Q_k) + \beta_{k+1} + \varepsilon \quad (3)$$

where logit function of variable x is $\ln(x/(1-x))$, β is the regression coefficient, and ε is an error term. Natural Language Processing (NLP) and semantic analysis are of paramount importance in such an attempt to filter ethnicity-related messages and identify relevant keywords of specific population subgroups for marketing segmentation.

Marketers can regard the sub-segments as the foothold for future growth (e.g., Raynor & Weinberg, 2004). The sub-segment of active social consumers can be a strong foothold for other sub-segments and other ethnic segments. It is because such social consumers "have the tools and the desire to share and socialize that information with those same trusted peers" and are willing to become "advocates" (Greenberg, 2010, p. 411). The conscripted foothold may be expanded and extended through online and offline consumers' networks. In ethnicity-based target marketing based on mixed business model, such an expectation is entirely legitimate because coherent ethnic groups tend to operate as both economic and social networks (Rangan, 2000). For instance, marketers can subsequently target existing social or economic networks via an initially conscripted online foothold. Pires et al. (2011) argue that "focusing on the measures that capture a group's social and business networks can enhance the formulation and effectiveness of marketing strategies" (p. 995). It may enhance the substantiality and reachability in the target marketing process by increasing engagement and activities. It also leads to a two-step approach both by multiplying a number of

touchpoints and intensifying them.

4. Evaluation

In order to assess how well the designed method supports a solution to the problem and demonstrates the use of the design scheme, we conducted a two-fold evaluation. A coverage test and an accuracy test of web demographics further ensured the quality of the digital marketing approach derived from web demographics as an alternative data source. An expert survey tested the degree of interest among marketing practitioners in the digital marketing solution.

4.1. Evaluating the quality of web demographics

To further ensure the quality of strategy program derived from web demographics, it is essential to evaluate the coverage, which refers to the sample size relative to the population. Indeed, coverage is an indication of sampling representativeness in statistics. In the context of target marketing, the coverage of both enumeration count and demographic attributes are important to support the formation of an informed marketing strategy. This study will use census data as a reference to verify the accuracy of web demographics and its coverage.

Because the acquisition of web demographics is made from people search engines, the coverage of web demographics depends on the comprehensiveness and representativeness of the data sources. This research adopted approximately 400 highly predictive surnames unique to Vietnamese from the Lauderdale and Kestenbaum (2000) and Falkenstein (2002) lists. While the volume of big data is growing exponentially with time, the exact coverage of web demographics relative to the documented population enumeration is unknown. As a pilot study that used 91 common Vietnamese surnames yielded about 40.3% and 10.4% from WhitePages and Intelius reported by the ACS 2009. With the addition of Addresses (www.addresses.com) and Zabasearch (www.zabasearch.com), a follow-up study in 2010 returned 81,440 (40.3%) and 101,061 (50%) records respectively (Intelius data was dropped because it returns only city location but not street address). However, some of these multi-sourced data are duplicated records of the same person. After a record linkage procedure that removes erroneous and duplicate subsets, the unique records represent a large sample of 194,861 persons (92.3%) using Census 2010 as a reference (Table 1).

Besides the total state-wide population count, the spatial distribution of population is compared at varying spatial scale. The Census Bureau, bounded by Title 13 of the U.S. code, is obligated to protect personal identity from any possible leakage. As a result, data masking practice was employed so that geographic units with a population less than a hundred would be changed to zero. A comparison of the VA between Census 2010 at the county level and web demographics revealed similar high VA populations around major metropolitan areas in Texas (Fig. 3). However, there were differences in the spatial pattern of VA count, especially in the rural counties where VA population is typically less than a hundred. The discrepancy between web and census enumeration becomes more evident as the spatial scale improves to the census tract level (Fig. 4). It is noted that the total VA population count

Table 1
Web demographics coverage of Vietnamese-Americans in Texas (adopted from Chow et al., 2012).

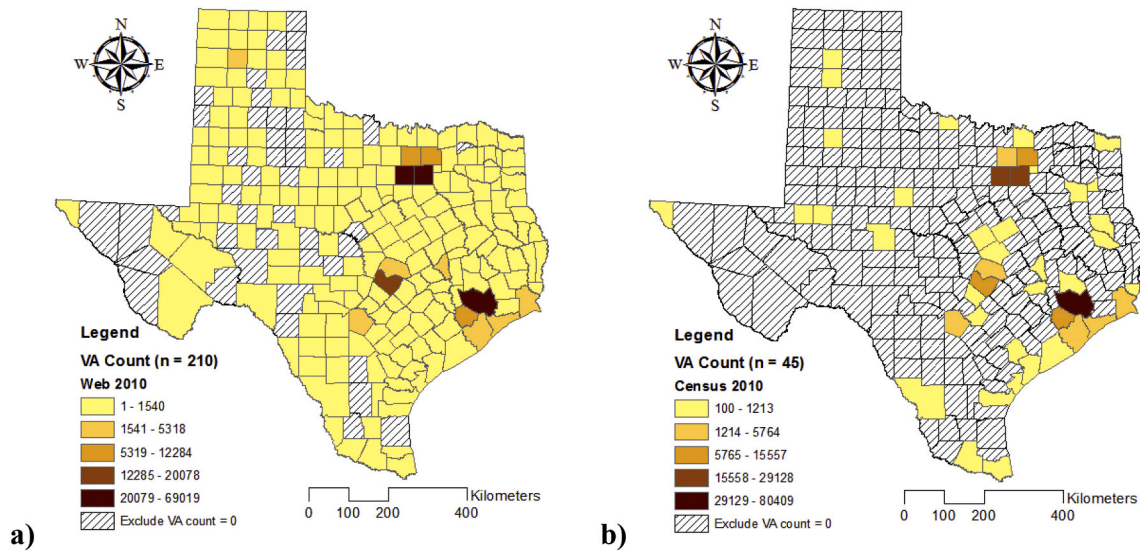|	Intelius	WhitePages	Addresses	Zabasearch	Total*
Raw data	20,957 (10.4%)	81,354 (40.3%)	81,440 (40.3%)	101,061 (50.0%)	263,855 (130.6%)
Valid records	18,737 (9.3%)	78,460 (38.8%)	79,673 (39.4%)	94,632 (46.8%)	252,765 (125.1%)
Unique records	–	53,681 (26.6%)	58,232 (28.8%)	76,712 (38.0%)	188,625 (93.4%)
Geocoded records	–	53,425 (26.4%)	58,245 (28.8%)	74,865 (37.1%)	186,535 (92.3%)

Fig. 3. A comparison of VA population count between web demographics (a) and Census 2010 (b) at the county level in Texas.
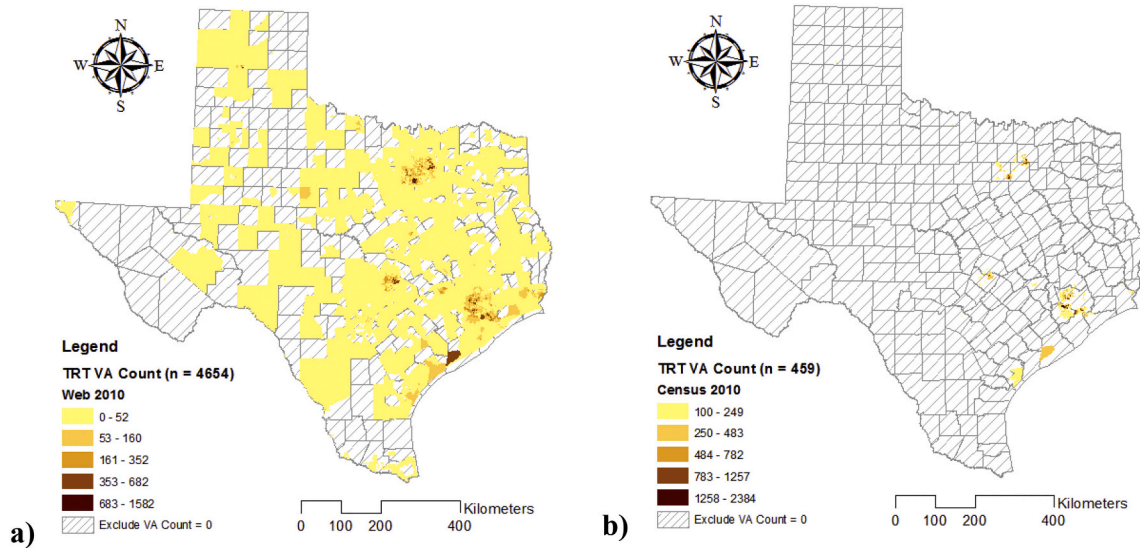

Fig. 4. A comparison of VA population count between web demographics (a) and Census 2010 (b) at the census tract level in Texas.

across Texas using census tracts from Census 2010 drops from 210,913 to 147,446, implying a serious undercount of the government data at that spatial scale.

Besides population count, it is also important that the demographic attributes (e.g., age) from web demographics are well-represented to support marketing segmentation. Some publicly available web demographics contain age data, which can be used to derive the age structure aggregated at any geographic units, ranging from census block to county/state (Fig. 2). While the age structure data from the Census Bureau is only tabulated by race (i.e., Asian) but not by ethnicity (i.e., Vietnamese), a preliminary comparison reveals interesting patterns at the census tract level (Fig. 5). Regardless of gender, the Asian age structure from Census 2010 and VA age structure from web demographics reveal similar decreasing trends in the age groups of 35 to 64 but a reverse, increasing trend in ages 65 and above. However, the age structures between 18 and 34 significantly differ between the Census 2010 and the web counterpart. This underestimation in the web records can be due to insufficient socio-economic activities among the younger individuals who have yet to leave enough digital trace (i.e., socioeconomic activities, utility subscription, etc.) on the Internet. The
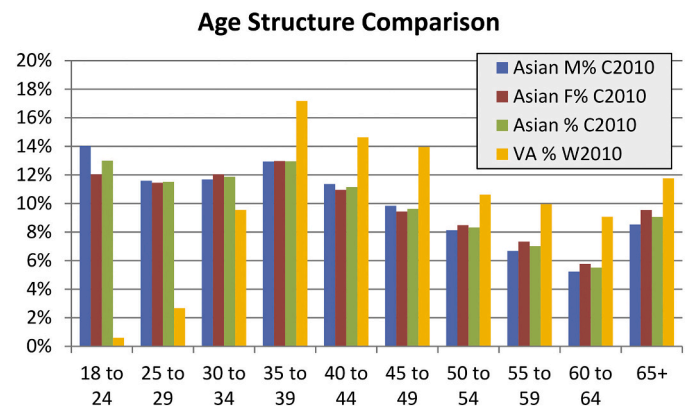

Fig. 5. A comparison of VA and Asian age structure between web demographics (W2010) and Census 2010 (C2010) at the census tract level in Texas.

underestimation of the age grou 18 to 34 in web demographics may present practical challenges and research opportunities for marketing segmentation in these population groups. Other sources of big data, such as social media, may be considered to fill this gap.

4.2. Expert test

We performed an expert survey to test whether there was a general interest among entrepreneurs in the developed process. Expert survey is done by presenting the problem and our process model as a solution to a bigger number of practitioners (Offermann et al., 2009). To evaluate the process model's viability, usability, usefulness, and business effect, we developed the following statements:

(1) the proposed method in the process model provides a viable solution to the mentioned problem in customer discovery and segmentation.
(2) in practice, the process model is usable.
(3) for practitioners, the process model is useful.
(4) the process model helps with improving business performance.

We presume that the design and development of the focal solution are supported if the survey supports the refined statements (Offermann et al., 2009).

The survey site was a metro-area of Austin, Texas. Through convenient sampling, we conducted interviews with 200 entrepreneurs that serve both businesses and consumers in a wide range of sizes. During the on-site interview, they observed the written description of the digital marketing solution. The interviewers also gave a detailed oral presentation to ensure the respondents fully understood the parameters. Notably, both presentations contained the three-fold issues and problems in the current practices of marketing segmentation, as stated earlier. Subsequently, the respondents answered a short list of questions, including four evaluation questions and descriptive information. We gained 141 usable surveys in total, and the response rate was 70.5%. The valid entries were yielded to analysis. Mean scores of annual sales, number of employees, and respondent's experience were 9 million, 85 employees, and 12 years, respectively.

When examining the results, we found that the respondents positively evaluated the described process model (*t*-values were all significant):

(1) 91.5% of the respondents (129 out of 141) answered that the presented method in the focal solution provides viable solutions to the mentioned problem in target marketing.
(2) 95.7% of the respondents (135 out of 141) answered that the focal solution is usable in practice.
(3) 71.6% of the respondents (101 out of 141) answered that the focal solution is useful for entrepreneurs.
(4) 72.3% of the respondents (102 out of 141) answered that the focal solution helps improve business performance.

A series of ANOVA tests revealed that their perceived importance of marketing, ethnic marketing, digital marketing, and business growth does not affect their evaluation ($p > .05$). This result proves that the focal solution has a general interest of entrepreneurs in practice.

5. Discussion and conclusions

Utilizing web demographics method, we proposed an approach of market-shaping target marketing for ethnic ventures' mixed business models. Web demographics can feasibly empower ethnic ventures' market-shaping strategy over mass customization and customerization (Wind & Rangaswamy, 2001). Adopting a design scientific approach, this study developed and tested a digital marketing procedure for ethnic ventures in a practical context. The orientation to design solution can

produce research output more relevantly (van Aken & Romme, 2009). Changes in the business environment and emergent perspectives, such as market-shaping in practice may require new approaches in designing and developing a marketing solution (e.g., Brown & Ellis, 2017).

5.1. Theoretical and practical implications

While the idea of shaping market has been around, more systematic inquiry on market-shaping is limited (Humphreys & Carpenter, 2018; Jaworski & Kohli, 2017; Nenonen et al., 2019), particularly for the concept of mixed business model (i.e., B2B2C). We believe that our design scientific approach can trigger future research to generate marketing artifacts to shape and manage the evolutions of markets. This study configured and re-configured a social segment in the market through the three stages of market-shaping target marketing. Our selection of ethnic ventures focusing on ethnic market provided an ideal context for the development of the market-shaping solution.

Traditional market segmentation approaches involve the division of a large heterogeneous market into several smaller homogeneous markets. The concept of market-shaping target marketing begins with an approachable foothold and facilitates a specific pathway of incremental targeting to market penetration. Ethnic ventures which engage in the market-shaping target marketing will be able to (1) create new opportunities to interlink resources of various members in the ethnic network, (2) integrate marketing activities being more able to combine strategic marketing process with the management of customer relationship, (3) "discover the value potential of linking intra- and inter-stakeholder resources in novel way", and (4) "trigger changes in various market characteristics to enable the formation of new resources linkages" (Nenonen et al., 2019, p. 618).

An important aspect of the efficient and effective management of a venture is to discover potential market segments to tailor market activities for faster growth (Ries, 2011). From web demographics, the primary attributes of individual records, such as full names and street addresses, provide the entry point for tailored market strategies. While it is relatively easy to identify gender from most names (especially English names), it is also possible to identify race and ethnicity by using surname analysis (Lauderdale & Kestenbaum, 2000). As discussed, household characteristics (e.g., single-family residence or multi-family dwelling) can be derived by geocoding the precise residential address and overlaying with an aerial photograph (Fig. 2c). Other auxiliary information in web demographics, such as age, phone number, and household size, can be used to conduct a preliminary customer profile in secondary clustering. In conjunction with the discount/reward card program common in grocery stores that tracks customers' transaction history, detailed customer behavioral patterns can be identified. At the neighborhood scale, the age structure can be generalized, as illustrated in Fig. 2d. The aggregated demographic and socioeconomic attributes from Census data can be used to infer other neighborhood characteristics, such as income level, education attainment, language preference, commuting pattern, etc. Marketers may tailor their strategic promotions based on neighborhood characteristics and potential needs. Other types of data from social media can also be combined to generate further insights and adaptable procedures.

Overall, our study revealed the importance of market-shaping strategy for ethnic ventures' mixed business models and the use of web demographics for the purpose. With enhanced cross-cultural experience, migrants and expatriates can better identify profitable opportunities and exhibit a higher level of entrepreneurial activity (Beckers & Blumberg, 2013; Vandor & Franke, 2016). On the other hand, the inability to make the connection between an opportunity in a dominant market and specific knowledge can make inexperienced, immigrant entrepreneurs focus specifically on their ethnic market, which is more familiar to them. The literature supports that human capital factors can determine the extent to which an entrepreneur recognizes opportunities (Gedajlovic et al., 2013; Light & Dana, 2013;

Venkataraman, 1997). The cognitive behaviors of entrepreneurs are bounded by their ability to gather appropriate information and the knowledge of how to process information (Woo et al., 1992). Our recommended web-demographic target marketing potentially integrates various types of market strategies into an adaptable, hybrid process due in part to the technological feasibility of big data analytics.

Personal information is huge and omnipresent on the Internet. Requiring only surname and references of location, people search engines can become the gateways to a wealth of demographic records. The resulting demographics of VA from WhitePages alone represented a sample size of about 40% of the estimated VA population according to the American Community Survey in 2009 (Chow, 2011, p. 488). This statistic symbolically reveals several advantages of web demographic data and its potential usefulness in the process of target marketing. With a clear-cut strategic direction, ethnic ventures can capitalize on the bountiful insights from big data by building up dynamic, appropriate, and substantial target marketing processes to develop adaptable market strategies.

5.2. Limitations and future studies

The focal digital marketing solution based on web demographics currently has several limitations that can be investigated in future research. Some strengths and limitations exist in all enumeration strategies, the web approach included. Some challenges in surname analysis include surnames common to multiple races/ethnicities (e.g., Lee), dummy and fictitious records, sampling bias towards economically active subgroups (i.e., younger children), married women and families with intermarriage. Future research can work to improve record linkage algorithms that incorporate cultural elements, information laws (e.g., Zipf's Law and Tobler's Law), and fusion with Census data to better adjust the information extracted from web demographics for marketing segmentation purposes. While data from web demographics cannot substitute the existing enumeration strategies, they do suggest a cost-effective and time-flexible alternative approach that can open a new venue for future research.

While the web demographics scrapped in this study are only limited to the primary attributes, linking the demographic attributes with socioeconomic attributes are common in geodemographic databases. Many data vendors have long been linking the demographic and socioeconomic datasets (e.g., credit card transactions) to conduct highly precise marketing segmentation and profiling. While these datasets have been well studied in marketing research, this study explores new and supplementary data sources that are more comprehensive in terms of sampling but also more readily available and usable in this interconnected cyberinfrastructure (e.g., social media).

Our approach to applying web demographics for ethnic ventures' mixed business models may potentially stimulate some future studies. More advanced applications can be designed for further marketing opportunities. Web demographics provide the names of individuals who may be associated with each record, which may include relatives, roommates, household members, and friends. Through connecting the associated persons, a social network of inter-related individuals can be identified. The myriad effect of "friends of friends" effectively reduces the degree of separation in sorting through intertwining human relationships. Daraghmi and Yuan (2014) suggest that the emergence of social media, dramatically reduced the degree of separation between any two people, even with the most unrelated features, from 6.0 to a range of 3.0–3.9. Hence, by matching and extending the network of associated people, who may also be connected with an unrelated person on the other end, web demographics can be used to construct a "smart network" for customer discovery and opportunity evaluation. Future studies would be plenteous as entrepreneurs diverge in their view of marketing opportunities (Gruber, Kim, & Brinckmann, 2015).

Ethics is a legitimate concern of linked geodemographic data, especially in this digital era. The ethical concerns include data privacy,

dataveillance, compliance, racial/ethnic inequality, etc. Who is collecting the data on whom, for what purpose, and how to keep them from the wrong hands or usage are active research agenda (Chow, 2013). Nevertheless, the web census approach can be used to enumerate a target population and to acquire and update other geographic features relevant to the business decision-making procedure (Chow et al., 2011). For example, using the Application Programming Interface (API) of Yelp, a list of other businesses and their corresponding reviews can be queried and retrieved to analyze the business landscape. Hence, the continual streaming of web demographics, VGI, and other big data can help in monitoring the customer and business landscapes that are important when devising appropriate entrepreneurial marketing strategies.

This study was based on the design scientific approach. We believe that design science research is one of the essential building blocks to fill up the theory-practice gap in the area of marketing. In a few business-related disciplines, the way to define and remedy the theory-practice gap has already been paraphrased with the emergence of design science research (e.g., Beloglazov et al., 2014; Huff, Tranfield, & Van Aken, 2006; Peffers et al., 2008). In marketing, the theory-practice gap commonly concerns a belief of many senior managers that "marketing is intrinsically art and experience, and is not amenable to the systematic approach" (Lilien, Rangaswamy, Van Bruggen, & Wierenga, 2002, p. 111). Although such a belief is somewhat reasonable and evidential, it will make more sophistication in making decisions become so arduous, and cause marketing management to be almost entirely dependent on individual talents (Lilien et al., 2002). It is increasingly believed that the recent development of design science is able to streamline various endeavors to fill up the gap in marketing and business research into a more solid foundation upon which researchers can gain more relevancy without losing rigor. Rigor and relevance are the main requirements of design science (Beck et al., 2013; Benbasat & Zmud, 1999; Hevner et al., 2004; Rosemann & Vessey, 2008). More design science research is very much needed in marketing.

Declaration of competing interest

None.

References

Aaker, J. L., & Maheswaran, D. (1997). The effects of cultural orientation on persuasion. *Journal of Consumer Research, 24*(3), 315–328.

van Aken, J. E., & Romme, A. G. L. (2009). Reinventing the future: Adding design science to the repertoire of organization and management studies. *Organization Management Journal, 6*, 5–12.

Beck, R., Weber, S., & Gregory, R. W. (2013). Theory-generating design science research. *Information Systems Frontiers, 15*(4), 637–651.

Beckers, P., & Blumberg, B. F. (2013). Immigrant entrepreneurship on the move: A longitudinal analysis of first-and second-generation immigrant entrepreneurship in the Netherlands. *Entrepreneurship and Regional Development, 25*(7–8), 654–691.

Beloglazov, A., Benergee, D., Hartman, A., & Buyya, R. (2014). Improving productivity in design and development of information technology (IT) service delivery simulation models. *Journal of Service Research, 18*(1), 75–89.

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems: The practice of relevance. *MIS Quarterly, 23*(1), 3–16.

Bird, M., & Wennberg, K. (2016). Why family matters: The impact of family resources on immigrant entrepreneurs' exit from entrepreneurship. *Journal of Business Venturing, 31*(6), 687–704.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*, 1–8.

Brotspies, H., & Weinstein, A. (2019). Rethinking business segmentation: A conceptual model and strategic insights. *Journal of Strategic Marketing, 27*(2), 164–176.

Brown, M., & Ellis, S. (2017). *Hacking growth: How Today's fastest-growing companies drive breakout success*. Random House.

Butler, J. S., & Greene, P. G. (1997). Ethnic entrepreneurship: The continuous rebirth of American enterprise. In D. L. Sexton & R. W. Smilor (Eds.), *The state of the art of entrepreneurship* (pp. 267–290). Chicago: Upstart.

Chaganti, R. R. S., Watts, A. D., Chaganti, R., & Zimmerman-Treichel, M. (2008). Ethnic-immigrants in founding teams: Effects on prospector strategy and performance in new internet ventures. *Journal of Business Venturing, 23*(1), 113–139.

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2013). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems, 63*, 67–80.

Chen, S., Borthwick, A., & Carvalho, V. R. (2011). The case for cost-sensitive and easy-to-interpret models in industrial record linkage. In *9th International workshop on quality in databases, august 29, 2011, Seattle, WA*.

Chow, T. E. (2011). Geography 2.0 – A Mashup perspective. In S. Dragicevic, & S. Li (Eds.), *Advances in Web-based GIS, mapping services and applications* (pp. 15–36). Taylor & Francis. Vol. of ISPRS Book Series.

Chow, T. E. (2013). We know who you are and we know where you live: A research agenda of web demographics. In D. Sui, S. Elmwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge. Volunteered geographic information (VGI) in theory and practice* (pp. 265–285). Springer.

Chow, T. E., Lin, Y., & Chan, W. D. (2011). The development of a web-based demographic data extraction tool for population monitoring. *Transactions in GIS, 15*(4), 479–494.

Chow, T. E., Lin, Y., Huynh, N. T., & Davis, J. (2012). Using web demographics to model population change of Vietnamese-Americans in Texas between 2000-2009. *GeoJournal., 77*(1), 119–134.

Cui, G., & Choudhury, P. (2002). Marketplace diversity and cost effective marketing strategies. *Journal of Consumer Marketing, 19*, 54–74.

Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the 1st workshop on social media analytics (SOMA)*. SOMA (2010).

Daraghmi, E. Y., & Yuan, S.-M. (2014). We are so close, less than 4 degrees separating you and me. *Computers in Human Behavior, 30*, 273–285.

Dibb, S., & Simkin, L. (2009). Implementation rules to bridge the theory/practice divide in market segmentation. *Journal of Marketing Management, 25*(3/4), 375–396.

Falkenstein, M. R. (2002). The Asian Pacific islander surname list: As developed from Census 2000. In *Proceedings of Joint Statistical Meeting 2002, New York, NY* (doi: 10.1.1.380.4707, Accessed 5 January 2021) https://citeseerx.ist.psu.edu/viewdoc/summary?.

Firat, A. F., & Shultz, C. J. (1997). From segmentation to fragmentation: Markets and marketing strategy in the postmodern era. *European Journal of Marketing, 31*(3/4), 183–207.

Gedajlovic, E., Honig, B., Moore, C. B., Payne, G. T., & Wright, M. (2013). Social capital and entrepreneurship: A schema and research agenda. *Entrepreneurship Theory and Practice, 37*(3), 455–478.

Greenberg, P. (2010). The impact of CRM 2.0 on customer insight. *Journal of Business & Industrial Management, 25*(6), 410–419.

Greene, P. (1997). Ethnic sponsorship of business activities: A resource based approach. *Journal of Small Business Management, 35*, 58–71.

Gruber, M., Kim, S. M., & Brinckmann, J. (2015). What is an attractive business opportunity? An empirical study of opportunity evaluation decisions by technologists, managers, and entrepreneurs. *Strategic Entrepreneurship Journal, 9*(3), 205–225.

Gulati, R., Nohria, N., & Zaheer, A. (2000). Strategic networks. *Strategic Management Journal, 21*(3), 203–215.

Harrison, D., & Kjellberg, H. (2010). Segmenting a market in the making: Industrial market segmentation as construction. *Industrial Marketing Management, 39*, 784–792.

Hassan, S., Craft, S., & Kortam, W. (2003). Understanding the new bases for global market segmentation. *Journal of Consumer Marketing, 20*(5), 446–462.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105.

Huff, A., Tranfield, D., & Van Aken, J. E. (2006). Management as a design science mindful of art and surprise. *Journal of Management Inquiry, 15*(4), 413–424.

Humphreys, A., & Carpenter, G. S. (2018). Status games: Market driving through social influence in the US wine industry. *Journal of Marketing, 82*, 141–159.

Iankova, S., Davies, I., Archer-Brown, C., Marder, B., & Yau, A. (2019). A comparison of social media marketing between B2B, B2C and mixed business models. *Industrial Marketing Management, 81*, 169–179.

Iansiti, M., & Levien, R. (2004). Strategy as ecology. *Harvard Business Review, 82*(3), 68–81.

Jaworski, B., Kohli, A. K., & Sahay, A. (2000). Market-driven versus driving markets. *Journal of the Academy of Marketing Science, 28*(1), 45–54.

Jaworski, B. J., & Kohli, A. K. (2017). Conducting field-based, discovery-oriented research: Lessons from our market orientation research experience. *Academy of Marketing Science Review, 7*(1–2), 4–12.

Jeon, J., & McSharry, P. (2013). The power of Twitter on predicting box office revenues. working paper. at http://users.ox.ac.uk/~mast1549/Jeon_BoxOfficeRevenueForecastingUsingTwitter.pdf.

Johanson, J., & Vahlne, J. E. (2011). Markets as networks: Implications for strategy-making. *Journal of the Academy of Marketing Science, 39*(4), 484–491.

Kjellberg, H., Azimont, F., & Reid, E. (2015). Market innovation processes: Balancing stability and change. *Industrial Marketing Management, 44*(1), 4–12.

Kjellberg, H., & Helgesson, C. F. (2006). Multiple versions of markets: Multiplicity and performativity in market practice. *Industrial Marketing Management, 35*(7), 839–855.

Kotler, P., & Keller, K. L. (2007). *A framework for marketing management*. Upper Saddle River, NJ: Pearson Education, Inc.

Lacoste, S. (2016). Perspectives on social media ant its use by key account managers. *Industrial Marketing Management, 54*, 33–43.

Lauderdale, D. S., & Kestenbaum, B. (2000). Asian American ethnic identification by surname. *Population Research and Policy Review, 19*, 283–300.

Lee, E., Fairhurst, A., & Dillard, S. (2001). Usefulness of ethnicity in international consumer marketing. *Journal of International Consumer Marketing, 14*(4), 25–48.

Leek, S., & Christodoulides, G. (2011). A literature review and future agenda for B2B branding: Challenges of branding in a B2B context. *Industrial Marketing Management, 40*(6), 830–837.

Leetaru, K. H., Wang, S., Cao, G., Padmanabhan, A., & Shook, E. (2013). Mapping the global twitter heartbeat: The geography of Twitter. *First Monday, 18*(5–6). May 2013. at http://journals.uic.edu/ojs/index.php/fm/article/view/4366/3654#p11 (accessed October 8, 2013).

Light, I., & Dana, L. P. (2013). Boundaries of social capital in entrepreneurship. *Entrepreneurship Theory and Practice, 37*(3), 603–624.

Lilien, G. L., Rangaswamy, A., & De Bruyn, A. (2007). *Principles of marketing engineering*. State College, PA: ISBM.

Lilien, G. L., Rangaswamy, A., Van Bruggen, G. H., & Wierenga, B. (2002). Bridging the marketing theory–practice gap with marketing engineering. *Journal of Business Research, 55*(2), 111–121.

Mandelbrot, B. (1966). Information theory and psycholinguistics: A theory of words frequencies. In P. Lazafeld, & N. Henry (Eds.), *Readings in mathematical social science*. Cambridge, MA: MIT Press.

Masurel, E., Nijkamp, P., & Vindigni, G. (2004). Breeding places for ethnic entrepreneurs: a comparative marketing approach. *Entrepreneurship & regional development, 16*(1), 77–86.

Mattsson, L. G., & Johanson, J. (2006). Discovering market networks. *European Journal of Marketing, 40*(3/4), 259–274.

McAfee, A., & Brynjolsfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 3–9.

Natter, M., Mild, A., Wagner, U., & Taudes, A. (2008). Planning new tariffs at tele.Ring: The application and impact of an integrated segmentation, targeting, and positioning tool. *Marketing Science, 27*(4), 600–609.

Ndofor, H. A., & Priem, R. L. (2011). Immigrant entrepreneurs, the ethnic enclave strategy, and venture performance. *Journal of Management, 37*(3), 790–818.

Nenonen, S., Storbacka, K., & Windahl, C. (2019). Capabilities for market-shaping: Triggering and facilitating increased value creation. *Journal of the Academy of Marketing Science, 47*(4), 617–639.

Neumeyer, X., Santos, S. C., Caetano, A., & Kalbfleisch, P. (2018). Entrepreneurship ecosystems and women entrepreneurs: A social capital and network approach. *Small Business Economics*, 1–15.

Offermann, P., Levina, O., Schönherr, M., & Bub, U. (2009). Outline of a design science research process. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology* (pp. 7–17).

Ojo, S., Nwankwo, S., & Gbadamosi, A. (2013). Ethnic entrepreneurship: the myths of informal and illegal enterprises in the UK. *Entrepreneurship & Regional Development, 25*(7-8), 587–611.

Peffers, K., Tuunance, T., Rothenberger, M. A., & Chatterjee, S. (2008). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77.

Penaloza, L. (1994). Atravesando fronteras/border crossings: A critical ethnographic exploration of the consumer acculturation of Mexican immigrants. *Journal of Consumer Research, 21*(June), 32–54.

Phinney, J. S. (1996). When we talk about American ethnic groups, what do we mean? *American Psychologist, 51*(9), 918–927.

Pires, G. D., Stanton, J., & Stanton, P. (2011). Revisiting the substantiality criterion: From ethnic marketing to market segmentation. *Journal of Business Research, 64*, 988–996.

Portes, A. (1981). 13 Modes of Structural Incorporation and Present Theories of Labor Immigration. *International Migration Review, 15*(1), 279–297.

Portes, A., & Zhou, M. (1992). Gaining the upper hand: Economic mobility among immigrant and domestic minorities. *Ethnic and Racial Studies, 15*, 491–522.

Quinn, L., Hines, T., & Bennison, D. (2007). Making sense of market segmentation: A fashion retailing case. *European Journal of Marketing, 41*, 439–465.

Ram, M., Jones, T., & Villares-Varela, M. (2017). Migrant entrepreneurship: Reflections on research and practice. *International Small Business Journal, 35*(1), 3–18.

Rangan, S. (2000). The problem of search and deliberation in economic action: When social networks really matter. *Academy of Management Review, 25*, 813–836.

Raynor, M. E., & Weinberg, H. S. (2004). Beyond segmentation. *Marketing Management, 13*(6), 22–28.

Ries, E. (2011). *The lean Startup: How Today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Books.

Rosemann, M., & Vessey, I. (2008). Toward improving the relevance of information systems research to practice: The role of applicability checks. *MIS Quarterly, 32*(1), 1–22.

Rossiter, J. R., & Chan, A. M. (1998). Ethnicity in business and consumer behavior. *Journal of Business Research, 42*(2), 127–134.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010*.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press.

Storbacka, K. (1997). Segmentation based on customer profitability: Retrospective analysis of retail bank customer bases. *Journal of Marketing Management, 13*, 479–492.

Triandis, H. C. (1996). The psychological measurement of cultural syndromes. *American Psychologist, 51*(4), 407–415.

U.S. Census Bureau. (2012). Most children younger than age 1 are minorities, Census Bureau reports. May 12, 2012. http://www.census.gov/newsroom/releases/archives/population/cb12-90.html (accessed on November 5, 2013).

Vandor, P., & Franke, N. (2016). See Paris and… found a business? The impact of cross-cultural experience on opportunity recognition capabilities. *Journal of Business Venturing, 31*(4), 388–407.

Vargo, S. L., Wieland, H., & Akaka, M. A. (2015). Innovation through institutionalization: A service ecosystems perspective. *Industrial Marketing Management, 44*, 63–72.

Venkataraman, S. (1997). The distinctive domain of entrepreneurship research. In J. A. Katz (Ed.), *3. Advances in entrepreneurship, firm emergence and growth* (pp. 139–202). Connecticut: JAI Press.

Venter, P., Wright, A., & Dibb, S. (2015). Performing market segmentation: A performative perspective. *Journal of Marketing Management, 31*(1–2), 62–83.

Wilson, K. L., & Portes, A. (1980). Immigrant enclaves: An analysis of the labor market experiences of Cubans in Miami. *American Journal of Sociology, 86*, 295–319.

Wind, J., & Rangaswamy, A. (2001). Customerization: The next revolution in mass communication. *Journal of Interactive Marketing, 15*(1), 13–32.

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of Marketing Research, 15*, 317–337.

Wind, Y., & Cardoza, R. (1974). Industrial market segmentation. *Industrial Marketing Management, 3*(2), 153–166.

Woo, C. Y., Folta, T., & Cooper, A. C. (1992). Entrepreneurial search: Alternative theories of behavior. In N. C. Churchill, S. Birley, W. D. Bygrave, D. F. Muzyka, C. Wahlbin, & W. E. Wetzel (Eds.), *Frontiers in entrepreneurship research 1992* (pp. 31–41). Babson College: Wellesley, Massachusetts.

Word, D. L., Coleman, C. D., Nunbziata, R., & Kominski, R. (2011). *Demographic aspects of surname from Census 2000, genealogy data: Frequent occurring surnames from Census 2000* (US Census Bureau).

Wyner, G. (2009). Beyond segmentation. *Marketing Management, 18*(2), 6–7.

Yankelovich, D., & Meer, D. (2006). Rediscovering market segmentation. *Harvard Business Review, 84*(2), 122–131.