EMERGING SCHOLARS

A Computational Study of *Xinghun* Marriage Between Young Gay People in China Based on Dating Profiles

ANNA DAVIDSEN BUHL

BO ÆRENLUND SØRENSEN

Abstract

This study examines the phenomenon of *xinghun*—a marriage between a homosexual man and a homosexual woman—in contemporary China. Through a quantitative analysis of more than 60,000 dating profiles from the website Chinagayles.com, we examine the motivations and concerns of young gay people in China who seek this type of marriage. The analysis shows that desire for financial and social stability is an important reason for many to engage in *xinghun*. Concern for parents and a wish to make them happy features as a major topic, and many explicitly search for a partner with a traditional gender expression to satisfy heteronormative expectations. However, the analysis also shows that *xinghun* cannot only be understood as a response to pressure; some seek *xinghun* for personal connection and friendship or as a means to gain more freedom and autonomy. Further, there are significant gender differences in approaches to *xinghun*, seen most clearly in the question of whether or not to have children. Finally, a central methodological contribution of this study is that the simple procedure of extracting n-grams from a large corpus of Chinese text can provide a powerful analytical method.

Keywords: *China*; *xinghun*; *homosexuality*; *dating profiles*; *family*; *n-grams*; *digital humanities*

Introduction

Homosexual relationships have become somewhat more accepted in Chinese society since the 1980s, but same-sex marriage is still not

officially recognised. Young gay people in China face the difficult challenge of reconciling their desires and identities with an intense pressure to conform to social norms and expectations. Some decide to marry a straight person without revealing their true desires, while others campaign openly for equality and inclusion at the risk of severe social consequences (Bao 2020). In recent decades, lesbians and gay men have increasingly turned to a different strategy: marrying each other. This practice is known as *xinghun* in Chinese. The term *xinghun* is a contraction of *xingshi hunyin*, literally meaning 'form marriage'. While it originated as a general term for people of all sexual orientations looking for non-traditional and/or sexless marriage, today it is primarily a term describing a marriage between a gay man and a lesbian woman (Huang 2018: 138). This version of *xinghun* first came to light in the late 1990s when a gay man put up an ad for a lesbian wife in the queer magazine *Pengyou Tongxin* (Huang 2018: 138). Since then, the number of *xinghun* marriages has increased and is estimated to have risen from 500 in 2004 to more than 100,000 in 2018 (Wang 2019: 15).

Dating websites have long since replaced newspaper ads when it comes to locating prospective partners. The largest website in China catering to individuals who are contemplating a *xinghun* marriage is Chinagayles.com, which has well over 400,000 user profiles and claims to have matched several tens of thousands of *xinghun* couples over the years.[1]

In this article, we take a quantitative approach to the study of *xinghun*. We utilise a simple yet powerful n-gram approach to find patterns in 62,710 dating profiles from Chinagayles.com, allowing us to gain a broad understanding of how the users present themselves and their wishes for a potential partner. The analysis sheds light on the specific motivations and concerns of those who seek a *xinghun* marriage in contemporary China, and, in particular, how they seek to reconcile social pressures with their individual identities and desires.

Xinghun marriage in context

In China, the social pressure to marry is considerable (Choi and Luo 2016; Engebretsen 2009; Kam 2013). Young people who do not marry at the age of their peers often find that their parents, relatives and wider social circles increasingly inquire about their status as singles, offer unsolicited advice or arrange blind dates (Kam 2013: 60). For

some gay people, *xinghun* serves as an effective way to physically and emotionally distance themselves from their parents (Huang 2018). Additionally, marriage may offer status and advancement in the social hierarchy (Wang 2019) or even improved career opportunities (Wang 2014). Based on a review of the academic literature, Song et al. (2023) note that motivations for choosing *xinghun* also include love, desire for children and family life, belief that marriage is the right thing to do, desire for companionship, religious beliefs and unwillingness to accept one's sexual orientation.

The opportunities, decisions and desires of individual marriage-seekers are also shaped by the larger societal context. In the case of China, this context has, to a significant extent, been shaped by policy choices made by China's Communist Party (CCP), which has both directly and indirectly catalysed significant changes in the institution of family after it assumed power in 1949. Starting with the 1950 Marriage Law, which was the first major revision of existing laws undertaken by the CCP (Altehenger 2018), the party-state's reorganisation of society has strengthened the nuclear family as a site for emotional sustenance and economic interdependence (Jankowiak and Moore 2016).

From the 1980s, privatisation and marketisation further increased the material importance of conjugal relationships by leaving families to bear the increasing costs of healthcare and old age pensions. The state further sought to privatise social reproduction costs through a new set of laws which placed considerable emphasis on the duty of children to provide material support for their parents, grandparents and younger siblings (Barbalet 2015). The current PRC law 'On Protection of the Rights and Interests of the Elderly' further stipulates that 'family members living apart from elderly persons shall frequently visit or send greetings to the elderly persons' (NPC Standing Committee 2018). In recent years, Chinese courts have become more responsive to faulting children for failing to live up to these regulations (Sangren 2021).

Increased life expectancy, low mandatory retirement age and high costs of childcare and private tuition in large cities with little subsidies for such services have also made the role of grandparents more prominent. This has created a situation in which many families have become more interdependent across generations, described as a tendency towards neofamilism (Yan 2018). For many young Chinese, this translates into a strong wish to live up to family expectations, as well as an intense pressure to have children to

continue the family line and sustain the family economically (Choi and Luo 2016).

Meanwhile, the state continues to maintain the household registration system as a means of administratively tying people to their place of residence, while also allowing the free movement of labour so as to spur the flourishing capitalist economy. This has created a situation in which between two and three hundred million labourers, predominantly of child-rearing age, make their living as migrant workers at a distance from their official place of residence (Zhong and Arnett 2018). For many, this provides opportunities for finding work in places far away from their parents, which may create a comfortable buffer from parental instruction and intrusion (Chang 2009).

In terms of sexuality, the Mao era (1949-1976) was generally a period of repression. State propaganda portrayed homosexuality alternately as a sickness requiring medical 'curing' and a moral flaw best overcome by socialism (Engebretsen 2009: 6). With China's political and economic reforms of the 1980s, along with the country's greater integration into the international community, homosexuality slowly became more accepted. Gay communities and events started to arise, especially in the largest cities and, with the rise of the internet, online (Bao 2018). Homosexuality was decriminalised in 1997 and taken off the list of mental illnesses in 2001.[2]

Nevertheless, Chinese gay people still face significant discrimination and social stigma. For instance, homosexuality is often associated with AIDS and pornography in mainstream media, generating hostility towards the gay community (Wang 2014: 14). This may be one reason why, according to a 2016 United Nations report on LGBTI experience in China, only around 15 per cent of Chinese gay people have disclosed their sexuality to their family (United Nations Development Programme 2016: 26). Furthermore, both marriage and adoption remain inaccessible for same-sex couples in China (Mendos *et al*. 2020: 327).

Despite these barriers, various forms of activism seeking to promote the rights and recognition of gay people have found expression throughout China for the past decades (Bao 2020; Engebretsen 2009). Sociologist Li Yinhe and others have campaigned for the legalisation of same-sex marriage since the early 2000s, but thus far to no avail (Bao 2020: 111). The government's approach to homosexuality has been summed up as the 'three "nots"': not

supporting, not opposing and not encouraging (Wang 2014: 14). However, several events in recent years indicate a less lenient attitude, including removal of numerous LGBT accounts on social media (Ren and Gui 2022: 116) and prohibitions of displays of homosexuality in TV shows and online audio-visual programmes (Mendos *et al.* 2020: 151).

Within homosexual communities in China, there are disagreements about *xinghun*, and some experience pressure to *not* engage in *xinghun,* but instead to 'come out' and be open about their sexuality (Huang 2018; Wang 2019). Some argue that *xinghun*, through perpetuating (heterosexual) marriage as a cultural imperative, is a way to silence homosexuality and make it invisible—and thus ultimately make it even harder for gay people in China to fight for rights and social existence (Choi and Luo 2016: 266). This leaves many Chinese gay people caught between pressure from two sides: conforming to public and parental expectations on the one hand, and claiming their identity and standing up for themselves as a marginalised group in Chinese society on the other.

Most existing studies of *xinghun* have been primarily qualitative, based on interviews or examining online discourses. Huang (2018) suggests that *xinghun* provides a way for Chinese gay people to reconcile their personal romantic desires with a wish to make their parents happy, challenging the notion that love and marriage need always go together. Choi and Luo (2016) focus more on the intense parental and societal pressure to marry. They consider *xinghun* a means of 'performing' a traditional heterosexual life, either for parents or together with parents towards the wider society. Conversely, Wang (2019) argues that *xinghun*, although shaped by heteronormativity and class inequality, cannot be reduced to a response to pressure, but serves as an active strategy to obtain more freedom and autonomy in a society where embracing an openly homosexual life may not be the most desirable option, even if Western queer discourses often assume it. Several studies emphasise that men and women often face different challenges and expectations when they engage in *xinghun* (e.g., Cho 2009; Engebretsen 2017; Gui and Meng 2023; Wang 2019).

A few studies have examined *xinghun* through dating profiles and online ads. Liu (2013) analyses 150 *xinghun* ads from the online community Tianya.cn through a qualitative theme analysis, manually identifying themes and patterns in the ads. Liu concludes that many

of the ads show an adherence to traditional Chinese values also present in heterosexual partner seeking, such as filial piety, classic gender roles and well-matched social status. She further finds that the ads serve as a space for *xinghun* seekers to negotiate between their public and private identities.

Gui and Meng (2023) analyse approximately 4,000 dating profiles from Chinagayles.com with a focus on gender roles and gendered performance. They conduct thematic coding, identifying a number of categories from a smaller sample of 200 profiles subsequently used to analyse the rest of the profiles. Combining this with a qualitative analysis, the authors conclude that *xinghun* upholds many of the heteronormative aspects underlying a conventional marriage, and that the expectations about *xinghun* held by many gay people are shaped by heteronormative discourse.

This study, being the first frequency-based and hitherto largest study of *xinghun* seekers, sheds new light on the motivations and concerns of young gay people in China seeking *xinghun*. By investigating term frequencies and user information in more than 60,000 dating profiles, we seek to not only identify common themes but show their prevalence and recurrence over a large number of *xinghun* seekers.

Method

For this study, we analysed 62,710 dating profiles from Chinagayles.com collected in November 2021. The profiles are freely available online and do not contain names, phone numbers or other personally identifiable information. We chose Chinagayles.com because it is the largest *xinghun* website in China and at the time of data collection had a website structure that made our analysis technically possible. Our main analysis is a quantitative frequency analysis coded in Python, but we also rely upon insights gained from reading through numerous profiles to provide context and avoid distancing ourselves from the individuals making up the collective aggregate.

Chinagayles.com has existed since 2005, which meant that taking a random sample of profiles might provide us with data stretching across more than fifteen years. While it would be interesting to investigate changes over time, doing so in practice would be very difficult, not only because the profiles do not contain information about their creation date, but also because users can update their profiles

at any time. However, since all profiles are numbered consecutively based upon the order in which they were created, we were able to limit our analysis to the most recent profiles and thus minimise the effects of changes over time. We collected the HTML code from all profiles starting from number 350,000 and continuing up until the most recent profile. After removing 5,133 deleted profiles, this left us with a total of 62,710 profiles.

Each profile contains four sections: personal information about the user (e.g. age, location, height, educational level); questions on various topics that the user can answer by selecting among predefined answers; specific partner requirements, such as location or minimum income level; and two free text fields, one labelled 'self-introduction' (*ziwo jieshao*), the other 'expectations for choosing a spouse' (*ze'ou qiwang*). Some users write a lot, others only a little. A small number of profiles contain only gibberish, but the scale of our analysis prevents these from impacting our results.

Information such as age, gender, location and answers to predefined questions is readily amenable to descriptive statistical analysis. Working with the two free text fields is considerably more challenging. When looking to count word frequencies, the most immediate issue is that written Chinese has no marker separating words, like a space in English (Jurafsky and Martin 2024: 19-20). Using dictionary-based word segmentation could pose problems for our corpus, since it mainly consists of informal text that does not necessarily follow standard vocabulary or grammar rules. Further, we would risk missing slang or *xinghun*-specific terms not listed in dictionaries.

Instead, we decided to use n-grams. An n-gram is a sequence of *n* neighbouring characters or words (Grimmer *et al.* 2022: 51). In essence, n-grams are based on the insight of linguist John Firth that '[y]ou shall know a word by the company it keeps' (Firth 1957: 11).[3] Our analysis uses n-grams at the character level and thus identifies which Chinese characters often appear next to each other. Utilising the Python Natural Language Toolkit (NLTK) Library (Bird *et al.* 2019), we made frequency distributions of the two, three and four characters most commonly occurring together—called *bigrams*, *trigrams* and *quadrigrams*, respectively. This approach relies exclusively on co-occurrence of characters and thus provides a convenient way of getting around the problem of identifying individual words. While this would not produce useful results on a small set of texts, it performs very well on a corpus of our size, where all the most

frequent bigrams found in the text are actual words, and only a small number of the most frequent trigrams and quadrigrams are not meaningful.

From our main corpus of all profile text, we created several sub-corpora. To examine differences in how users referred to themselves versus a potential partner, we divided the text based on whether it was written in the 'self-introduction' or the 'partner expectations' field. In 2,750 profiles, the text of the two fields were identical. In those cases, we treated the 'partner' space as empty to avoid counting the same text twice. Further, we separated the text by gender, creating sub-corpora of male and female users. This resulted in nine total corpora, as shown in Table 1. Before performing frequency analysis, we cleaned the profile texts by removing punctuation and numbers (Grimmer *et al.* 2022: 52).

Table 1: Overview of corpora.

Corpus		Total # of profiles	Total character tokens before cleanup	Total character tokens after cleanup
All	All text, all users	62,710	6,500,828	5,483,281
	'Self' field, all users		4,094,613	3,448,765
	'Partner' field, all users		2,406,215	2,034,516
Male	All text, male users	43,747	4,571,070	3,857,897
	'Self' field, male users		2,930,781	2,467,173
	'Partner' field, male users		1,640,289	1,390,724
Female	All text, female users	18,963	1,929,758	1,625,384
	'Self' field, female users		1,163,832	981,592
	'Partner' field, female users		765,926	643,792

One issue, which revealed itself after reading through a great number of individual profiles, is that many users do not separate neatly between what they input in the 'self' and 'partner' text fields. Some write about themselves in the field devoted to describing their desired partner, while others write about their desired partner in the

self-description field. Attempting to clean up this 'messiness' of the data would not only be exceedingly difficult to do programmatically, it would also entail altering the data in unpredictable ways. Therefore, we primarily analysed the two text fields taken together, emphasizing differences between 'self' and 'partner' frequencies only in a few cases where the difference was relevant and remarkable. This issue is a good example of why it is important to conduct a thorough exploration of data ahead of analysing it computationally.

In our final list of the most frequent n-grams, the only adjustment we have made is to remove trigrams that were clearly part of the same, equally frequent quadrigram. For example, *gongzuowen* and *zuowending* are clearly non-meaningful partitions of the four-character phrase *gongzuo wending* ('job stability'). This highlights one weakness of the n-gram approach: the potential for words or phrases to be cut off or represented in an incomplete form, or to be embedded in longer n-grams. On the other hand, n-grams effectively reveal common trends and tendencies across the data without relying on considerations about what to count as a word or which words to look for. Even good algorithm-based segmentation tools for Chinese, such as Jieba[4], make mistakes, overlook words or group characters incorrectly together. Our approach provides a straightforward way of identifying both the most frequent words, idioms and collocations in a text corpus without relying on algorithmic decisions and capabilities.

A frequency approach provides little insight into individual users and uncommon uses of language, but it is an efficient means of identifying what users most regularly bring up. As noted by Hart *et al.* (2005: 248): 'Necessarily, counting things has distanced us from the texts being analysed [...] but it lets us see patterns that would have been missed close up'. This frequency-based approach to analysing language has proven itself surprisingly useful, as discussed at length by Underwood (2019). The simplicity of the n-gram approach, we argue, is a value in itself, because its results are easier to understand and interpret than those generated by more complicated algorithms.

Results

User demographics

First, let us explore the demographic composition of users on Chinagayles.com (Figure 1). Out of the 62,710 users included in the analysis, 69.8 per cent are male and 30.2 per cent female (no other

options are available on the site). 70.2 per cent of users are between 25 and 34 years of age, and female users are on average slightly younger than male users.
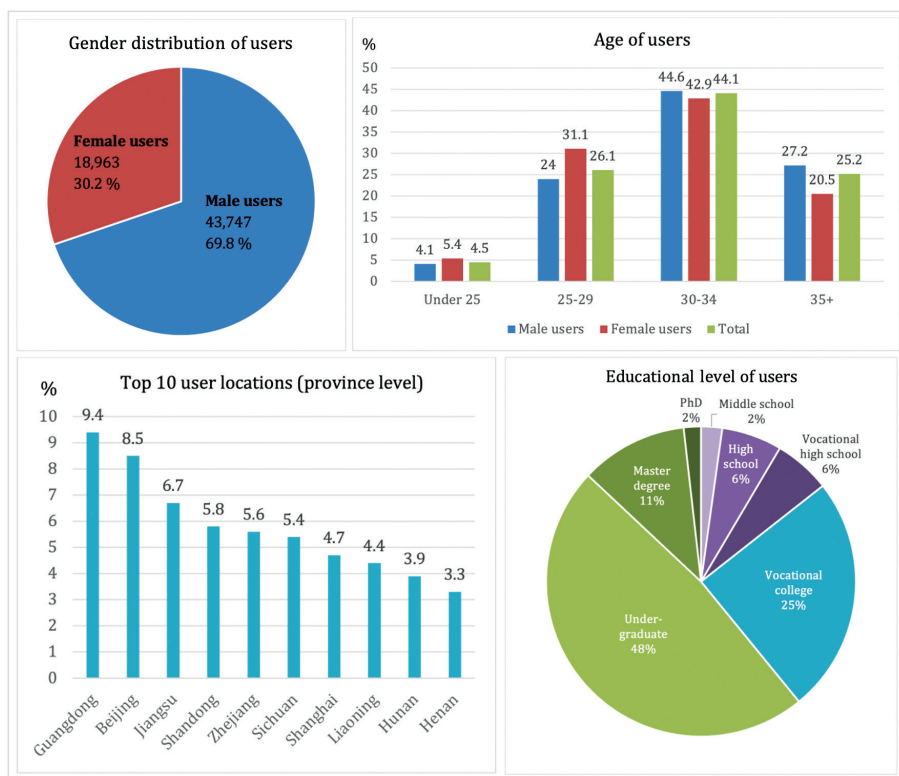

Figure 1: User demographics.

China's largest cities are significantly over-represented on the site, with Beijing and Shanghai being home to 8.5 and 4.7 per cent of users, respectively, while only making up 1.5 and 1.8 per cent of China's total population (National Bureau of Statistics of China 2022). Furthermore, the users are generally well-educated with 60.9 per cent of the users who disclosed their education level holding a university degree.

Overall, what stands out the most at the demographic level is the predominance of male users. This also matches the finding of Liu (2013) among *xinghun* ads on Tianya.cn. Part of the reason may be that, as Jankowiak and Moore (2016: 87) note, *xinghun* can be particularly psychologically demanding for women, who are often expected to take primary responsibility for family events and emotional contact

with parents, in-laws, cousins, etc. Another reason may be that the homosexual stigma generally falls more heavily on men than on women in China (Jankowiak and Moore 2016: 86).

Most frequent n-grams

Tables 2 and 3 present the 25 most frequent bigrams, trigrams and quadrigrams across all profile text. From this, we can get an idea about common themes and considerations for users on Chinagayles.com.

Table 2: Most frequent bigrams and trigrams in all profile text.

(* = truncated)

TOP 25 BIGRAMS			TOP 25 TRIGRAMS		
1.	*xiwang* 希望 (hope)	49,369	1.	*yao xiaohai* 要小孩 (want children)	10,245
2.	*keyi* 可以 (be able to)	40,035	2.	*you wending* 有稳定 (have a stable …)	9,529
3.	*gongzuo* 工作 (work)	32,658	3.	*yao haizi* 要孩子 (want children)	8,971
4.	*fumu* 父母 (parents)	32,656	4.	*bu hun quan* 不混圈 (non-chaotic social circles)	8,751
5.	*xinghun* 形婚 (xinghun)	28,040	5.	*zhao yi ge* 找一个 (look for a)	7,416
6.	*wending* 稳定 (stable)	26,645	6.	*xiwang ni* 希望你 (hope you)	7,288
7.	*shenghuo* 生活 (life)	25,926	7.	*xiwang zhao* 希望找 (hope to find)	6,577
8.	*yi ge* 一个 (one)	22,047	8.	*xiwang neng* 希望能 (hope to be able to)	6,174
9.	*haizi* 孩子 (children)	21,452	9.	*you guding* 有固定 (have a regular …)	6,154
10.	*xingge* 性格 (personality)	20,614	10.	*nüxinghua* 女性化 (feminine)	5,153
11.	*pengyou* 朋友 (friend)	18,190	11.	*ziji de* 自己的 (one's own)	4,995
12.	*benren* 本人 (I)	17,356	12.	*de shenghuo* 的生活 (… life)	4,432
13.	*xiaohai* 小孩 (children)	16,921	13.	*bu ganshe* 不干涉 (not interfere)	3,963
14.	*yiqi* 一起 (together)	16,302	14.	*ren yiyang* 人一样 (like …)	3,816
15.	*jingji* 经济 (economy)	15,446	15.	*zai yiqi* 在一起 (together)	3,794

	TOP 25 BIGRAMS			TOP 25 TRIGRAMS	
16.	*duli* 独立 (independence)	15,299	16.	*hao xiangchu* 好相处 (get along well)	3,459
17.	*shanliang* 善良 (kind-hearted)	14,929	17.	*dou keyi* 都可以 (can both)	3,313
18.	*duifang* 对方 (the other party)	14,646	18.	*xingge hao* 性格好 (good personality)	3,311
19.	*huxiang* 互相 (each other)	13,838	19.	*ding banlü* 定伴侣 (stable/regular partner *)	3,276
20.	*jiating* 家庭 (family)	13,645	20.	*wending de* 稳定的 (stable)	3,142
21.	*buyao* 不要 (do not want)	13,531	21.	*zuihao shi* 最好是 (it would be best)	3,120
22.	*zuihao* 最好 (it would be best)	13,442	22.	*xiang zhao yi* 想找一 (wish to find a)	3,015
23.	*ziji* 自己 (oneself)	13,196	23.	*you ziji* 有自己 (have one's own)	2,989
24.	*xiangchu* 相处 (get along)	13,028	24.	*heshi de* 合适的 (suitable)	2,988
25.	*hunhou* 婚后 (after marriage)	12,062	25.	*buyao tai* 不要太 (do not want too …)	2,966

Table 3: Most frequent quadrigrams in all profile text.

(* = truncated)

	TOP 25 QUADRIGRAMS	
1.	*buliang shihao* 不良嗜好 (bad habits)	9,904
2.	*jingji duli* 经济独立 (economic independence)	9,824
3.	*shenti jiankang* 身体健康 (healthy body)	9,445
4.	*wu buliang shi* 无不良嗜 (not have bad habits *)	8,582
5.	*gongzuo wending* 工作稳定 (job stability)	8,239
6.	*wending gongzuo* 稳定工作 (stable job)	5,077
7.	*xiwang duifang* 希望对方 (hope the other party)	5,072
8.	*you zerenxin* 有责任心 (be responsible)	4,980
9.	*xinghun duixiang* 形婚对象 (xinghun partner)	4,360
10.	*xingge kailang* 性格开朗 (bright and cheerful)	4,153
11.	*xiaoshun fumu* 孝顺父母 (filial piety)	3,977
12.	*you wending gong* 有稳定工 (have a stable job *)	3,929

TOP 25 QUADRIGRAMS (continued)		
13.	*wuguanduanzheng* 五官端正 (regular features)	3,496
14.	*tongqingdali* 通情达理 (sensible and reasonable)	3,247
15.	*shuangfang fumu* 双方父母 (both parties' parents)	3,051
16.	*xiwang zhao yi* 希望找一 (hope to find a)	2,875
17.	*yiyang xiangchu* 一样相处 (get along like …)	2,780
18.	*buyao haizi* 不要孩子 (do not want children)	2,626
19.	*buyao xiaohai* 不要小孩 (do not want children)	2,544
20.	*wending shouru* 稳定收入 (stable income)	2,459
21.	*wang zhao yi ge* 望找一个 (hope to find a *)	2,425
22.	*qinren yiyang* 亲人一样 (like family*)	2,417
23.	*huxiang bangzhu* 互相帮助 (help each other)	2,406
24.	*wending jingji* 稳定经济 (stable economy)	2,369
25.	*guding banlü* 固定伴侣 (regular partner)	2,339

All values in the above tables and in frequency tables throughout the article are absolute frequencies. Since the same word can appear multiple times within one profile, we do not give a percentage of users who wrote a given n-gram in their profile. Instead, tables 4-11 include a proportion of how many times the n-gram appears per 100 profiles.

In the following sections, we dive deeper into what the frequencies can tell us about the users' motivations to engage in *xinghun*. We have divided our analytical findings into three clusters: 1) Material issues and considerations, such as income and car ownership; 2) Personal characteristics, such as appearance and morality; 3) Family issues and considerations, including how to deal with parents and the issue of childbearing. As will become clear, the majority of highly frequent n-grams speak to one or more of these categories.

Material issues and considerations

Financial expectations lay heavily on young Chinese—especially men—who look to marry (Osburg 2024; Yu and Xie 2015). Economic independence may be particularly important for gay people, both as a means to acquire their own place to live and reduce parental influence (Wang 2014), and as a way to shore up individual social status that may be threatened by not conforming to traditional family or gender norms (Engebretsen 2017). In general, money allows you freedom to shape your own life and grants you certain opportunities

that may make it easier to handle social expectations. Put simply, you can 'afford not to care' about other people's judgment (Engebretsen 2009: 11). According to Liu and Tan (2020), China's neoliberal reforms in recent decades have made economic considerations primary for many *xinghun* seekers.

Viewed against this background, it is not surprising that we see several terms related to finances at the top of the frequency lists. The word 'job' (*gongzuo*) is the third most frequent bigram, 'economy' (*jingji*) is 15th, and 'economic independence' (*jingji duli*) is the second most frequent quadrigram. What is emphasised again and again is not the importance of a *well-paid* job, but a *stable* job. This is reflected in the quadrigram frequencies, which shows that 'job stability' (*gongzuo wending*) and 'stable job' (*wending gongzuo*) are fifth and sixth, and that 'stable income' (*wending shouru*) and 'stable economy' (*wending jingji*) are in places 20 and 24. Further, only 13.6 per cent of users specify a minimum requirement for their partner's income, which underlines the prioritising of economic stability over a specific income level.

Generally, stability features as a central concern in the profiles. The term 'stable' (*wending*) on its own is the sixth most frequent bigram, and with 'have' in front (*you wending*), it is the second most frequent trigram. Additionally, we see terms like 'regular partner' (*guding banlü*) and 'non-chaotic social circles' (*bu hun quan*), suggesting that the wish for stability is not only financial but social as well.

Another frequent term related to economic status is the phrase 'have an apartment and a car' (*youfangyouche*, or in reverse order, *youcheyoufang*), mentioned 3,321 times. This is a set phrase often used to discuss the prerequisites for entering into marriage in contemporary China (Osburg 2024: 92). As Liu (2013: 502) notes, apartment ownership may be especially important in *xinghun* marriages because of the extra practical dimensions involved in potentially having multiple partners in the picture, and it can be an important way for the married parties to maintain independence, not only from their parents but from each other as well.

Table 4: Frequency of 'have an apartment and a car'.

Frequency table: 'have an apartment and a car' (*youfangyouche* 有房有车 and *youcheyoufang* 有车有房, combined)			
	Male users	Female users	All users
'Self' field	2,261 (5.2 times/100 profiles)	623 (3.3 times/100 profiles)	2,884 (4.6 times/100 profiles)
'Partner' field	86 (0.2 times/100 profiles)	351 (1.8 times/100 profiles)	437 (0.7 times/100 profiles)
Total	**2,347 (5.4 times/100 profiles)**	**974 (5.1 times/100 profiles)**	**3,321 (5.3 times/100 profiles)**

The frequency of 'have an apartment and a car' shows a distinct gender difference (Table 4). Female users mention it much more in their 'partner' field than male users, who primarily use the phrase when writing about themselves. This fits with the notion that men in China generally face greater pressure to be breadwinners (Yu and Xie 2015). Chen (2009) observes the same in a *xinghun* context and notes that many gay men seeking *xinghun* feel that lesbians make high financial demands on them. This is further emphasised by several male users who put forth economic conditions, such as one 32-year-old who writes: 'You have to give back the bride money I pay to you, and for economic matters, we split evenly'.

Personal qualities

Several terms related to character and personal qualities are used frequently in the profiles. 'Personality' (*xingge*) is the 10th most frequent bigram, while 'good personality' (*xingge hao*) is 18th in the list of trigrams, and 'bright and cheerful' (*xingge kailang*) is 10th in the list of quadrigrams. Relatedly, 'kind-hearted' (*shanliang*) is the 17th most frequent bigram. More specific wishes for personal traits are also present; for example, many users highlight the importance of being 'sensible and reasonable' (*tongqingdali*) and 'responsible' (*you zerenxin*), which are both in the top 15 quadrigrams.

Another topic of importance is the physical appearance of the potential *xinghun* partner. The two terms *zhangxiang* and *waibiao*, which can both be translated as 'appearance', are mentioned a total of

7,932 times, almost equally between male and female users (Table 5). At 13th place in the list of quadrigrams, we find *wuguanduanzheng*, an idiom meaning to have a regular, well-proportioned face.

Users comment both on their own appearance and their wishes for that of a potential partner. For example, a 30-year-old male user writes: 'I look quite ok, and I hope to find a relatively good-looking *xinghun* girlfriend who lives in Chengdu'. Many users mention their height and weight, and several have specific requests for the minimum or maximum height of their potential partner.

Table 5: Frequency of 'appearance'.

Frequency table: 'appearance' (*zhangxiang* 长相 and *waibiao* 外表, combined)			
	Male users	Female users	All users
'Self' field	3,170 (7.3 times/100 profiles)	1,356 (7.2 times/100 profiles)	4,526 (7.2 times/100 profiles)
'Partner' field	2,550 (5.9 times/100 profiles)	856 (4.5 times/100 profiles)	3,406 (5.5 times/100 profiles)
Total	**5,720 (13.1 times/100 profiles)**	**2,212 (11.7 times/100 profiles)**	**7,932 (12.6 times/100 profiles)**

An interesting word when talking about appearance is 'feminine' (*nüxinghua*).[5] It appears 5,153 times in total and is the 10th most popular trigram, but its distribution is far from equal between genders (Table 6). In male profiles, it appears 10.1 times per hundred profiles, compared to 3.9 times for female ones. Additionally, we can see that male users are more likely to mention it in the field for describing their partner, while female users overwhelmingly use it in the field to describe themselves. Other terms denoting specific aspects of traditional femininity are also present, such as 'long hair' (*changfa*), which is used more by the female users, primarily in the 'self' field (Table 7).

Table 6: Frequency of 'feminine'.

Frequency table: 'feminine' (*nüxinghua* 女性化)			
	Male users	Female users	All users
'Self' field	1,817 (4.2 times/100 profiles)	594 (3.1 times/100 profiles)	2,411 (3.8 times/100 profiles)
'Partner' field	2,590 (5.9 times/100 profiles)	152 (0.8 times/100 profiles)	2,742 (4.4 times/100 profiles)
Total	**4,407 (10.1 times/100 profiles)**	**746 (3.9 times/100 profiles)**	**5,153 (8.2 times/100 profiles)**

Table 7: Frequency of 'long hair'.

Frequency table: 'long hair' (*changfa* 长发)			
	Male users	Female users	All users
'Self' field	212 (0.5 times/100 profiles)	976 (5.1 times/100 profiles)	1,188 (1.9 times/100 profiles)
'Partner' field	715 (1.6 times/100 profiles)	27 (0.1 times/100 profiles)	742 (1.2 times/100 profiles)
Total	**927 (2.1 times/100 profiles)**	**1,003 (5.3 times/100 profiles)**	**1,930 (3.1 times/100 profiles)**

Many users comment on this wish for femininity. A 32-year-old male user writes: 'I am very "man", quite straight looking, so I'm looking for a womanly-leaning [*pian nüren*] lesbian (you shouldn't be too manly)'. Another male user, 26 years old, describes his potential partner with only one sentence: 'As long as you're a feminine-leaning [*pian nüxinghua*] girl, it's fine'. Similar cases can be found among female users who do *not* want their future spouse to be woman-like. For example, a 33-year-old female user writes: 'My appearance is not bad. My requirement is that my partner is not womanly [*bu niang*]'.

Engebretsen (2009: 9) describes this focus on femininity versus masculinity in gay communities as a way of 'attaining a balance between masculine and feminine gendered qualities in appearance, personality, and social and sexual behaviour'. These wishes tie in with

deeply rooted gender norms and expectations about a man's and a woman's different roles in a successful marriage. As Chen (2009: 116) notes: 'Even though it is a *xinghun* marriage, it is still founded on the basis of a traditional marriage'. Such wishes for a traditionally feminine lesbian partner may be an unconscious reproduction of heteronormative narratives about what makes a good wife and marriage (Cho 2009; Engebretsen 2017).

The preference for traditional gender expression can also be seen as a more conscious effort to create an image of conformity to the outside world (Gui and Meng 2023). This is closely related to Choi and Luo's (2016) framework of performing the family, and *xinghun* as an acting out of conventional family ideals. If the goal is to showcase a heterosexual marriage and maybe even persuade your family that you are heterosexual, fulfilling stereotypical gender roles is a way to decrease the likelihood that you will stand out. As Cho (2009: 415) notes, *xinghun* 'requires convincing performances in order to be effective'. Engebretsen similarly finds it a common belief that 'a particularly effeminate gay man or masculine lesbian [is] more likely to cause suspicion and problems with family and colleagues', whereas *xinghun* partners who conform to gender expectations are usually 'considered less risky (more likely to pass), more trustworthy, and capable of acting their prescribed *xinghun* role' (Engebretsen 2017: 172).

Aside from their partner's gender expression, health and habits are also important factors for many users. At the very top of the list of quadrigrams, we find both 'healthy body' (*shenti jiankang*) and 'bad habits' (*buliang shihao*), which is the most frequent of all quadrigrams in the profiles. The latter is usually used in the negative, 'not have bad habits' (*wu buliang shihao*), which shows up in truncated form (*wu buliang shi*) as the fourth most frequent quadrigram. These 'bad habits' are used liberally by the users as an overarching term to cover a range of undesired behaviour such as smoking, drinking and other health choices as well as moral issues such as gambling or promiscuity. As can be seen in Figures 2 and 3, users indicate little interest in drinking and smoking, but since less than one in four have answered these questions, we cannot draw firm conclusions.
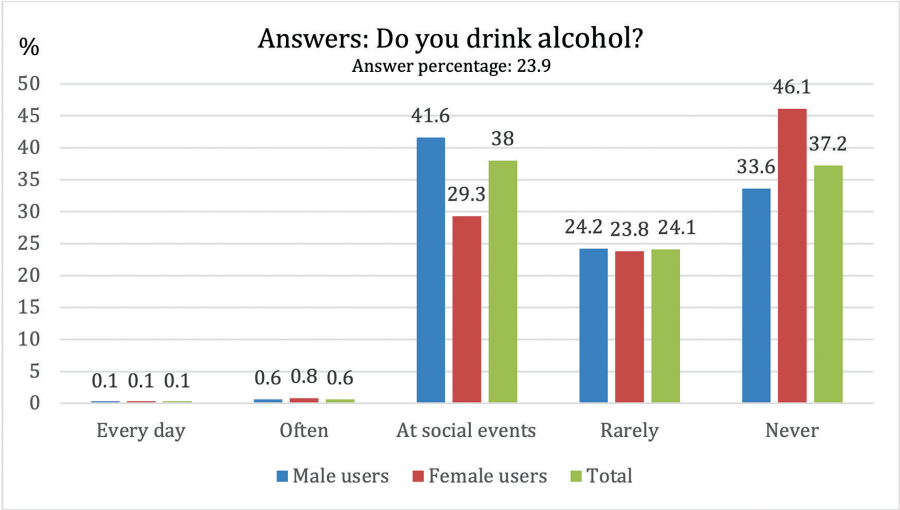
Answers: Do you drink alcohol?
Answer percentage: 23.9

%

	Every day	Often	At social events	Rarely	Never
Male users	0.1	0.6	41.6	24.2	33.6
Female users	0.1	0.8	29.3	23.8	46.1
Total	0.1	0.6	38	24.1	37.2

Figure 2: Users' alcohol habits.

Answers: Do you smoke?
Answer percentage: 23.8

%

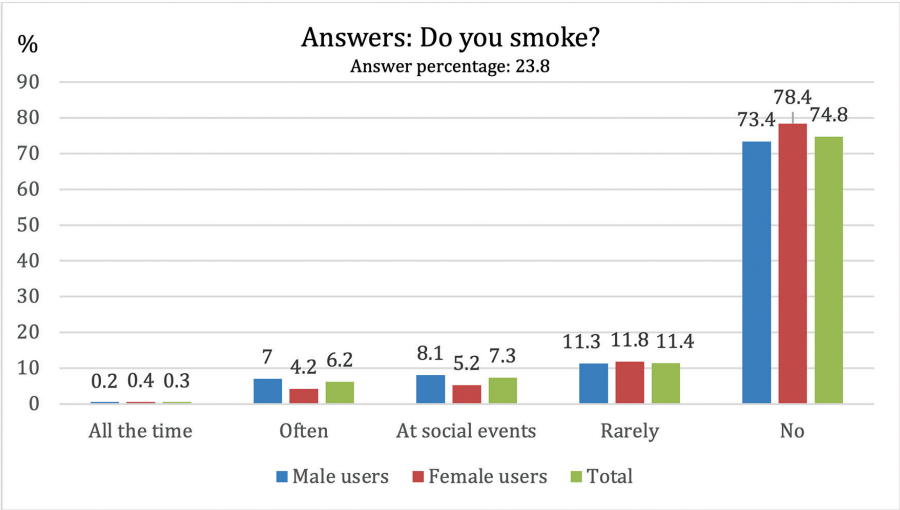|	All the time	Often	At social events	Rarely	No
Male users	0.2	7	8.1	11.3	73.4
Female users	0.4	4.2	5.2	11.8	78.4
Total	0.3	6.2	7.3	11.4	74.8

Figure 3: Users' smoking habits.

Another interesting concept related to habits, mentioned briefly above, is *bu hun quan*, which might be best translated as 'non-chaotic social circles' (literally 'not confuse circles'). This phrase is not found in standard dictionaries, but it is the fourth most common trigram in the profiles—a good example of a word that might be missed if using a dictionary-based word segmentation algorithm. The phrase

bu hun quan means that one does not associate with people who engage in 'undesirable' acts such as the bad habits discussed above. Its opposite would refer to someone who lacks moral judgment or chooses friends carelessly. The need to stress that one does not behave improperly may reveal a self-stigmatising attitude among users on the site. Engebretsen (2009) argues that homosexual relationships in China are often viewed as inherently 'chaotic' and 'turbulent', even by gay people themselves. As she notes: 'being lala [lesbian] and participating in lala sociality are in crucial ways evidence of being abnormal, unstable, turbulent, and chaotic, versus the desirable opposites of normal, stable, and harmonious—or, as many of the women I spoke with put it, being "just like everyone else"' (Engebretsen 2009: 9). Liu (2013: 503) makes a similar point when she notes that many users on Tianya.cn use *bu hun quan* as a way to distance themselves from a 'stereotypical gay lifestyle', and that this act of distancing is generally advertised as a desirable quality.

At the same time, the explicit wish for a partner who is 'not chaotic' further underlines the importance of stability for many gay people seeking *xinghun*. A *xinghun* marriage can be a way to achieve a sense of security and 'normality' in a life otherwise deemed 'chaotic'.

Children, family and married life

Deciding whether to have children is a topic of high concern to the users of Chinagayles.com. If we add up the two most common ways of referring to children, *haizi* and *xiaohai*, we find that children are referenced 38,373 times, or 61.2 times per 100 profiles. Since marriage is generally regarded as the only acceptable setting to foster children, and non-married couples have access to neither adoption nor assisted reproductive technologies, getting married is often the only option for gay people in China who wish to procreate (Gui and Meng 2023: 6).

In total, 42.5 per cent of users have answered the site's pre-defined question of whether they want children (Figure 4). In this aspect, we see a clear gender difference. 34.6 per cent of the female users who answered the question state that they do not want children, while 24.6 per cent do want children. In contrast, half of the male respondents say they want children, while only 13 per cent do not.
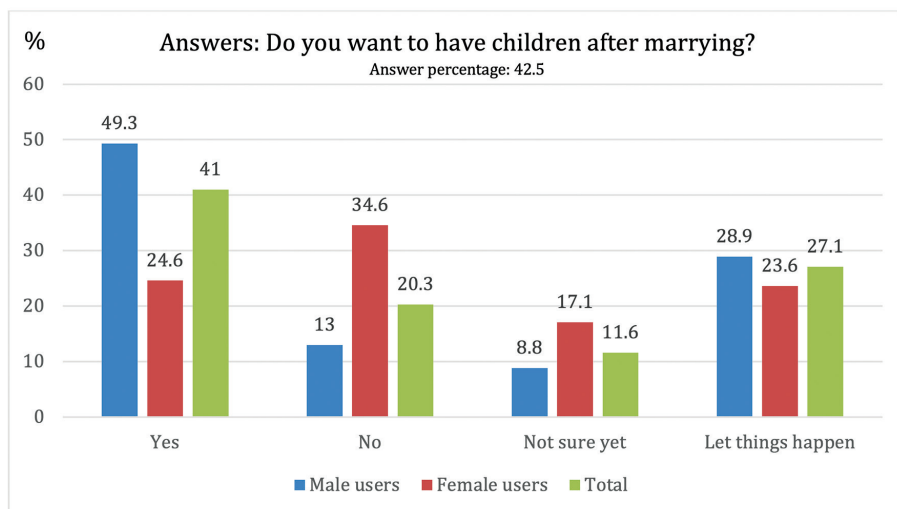
Figure 4: Users' attitude towards having children.

Those results match the findings of Choi and Luo (2016), who argue that gay men and lesbians face different types of pressure when it comes to entering marriage. While fathering children is a strong social duty for men, a woman's parents are often more concerned about their daughter finding a husband to take care of her. This creates a discrepancy in how gay men and lesbians view the question of children, and many lesbian women may feel a need to protect their sexual boundaries (Choi and Luo 2016).

The importance attributed to the topic of children is also reflected in the fact that two phrases meaning 'want children' (*yao haizi* and *yao xiaohai*) are the most and third most frequent of all trigrams. Together, they appear more than 19,000 times. This, however, might give a wrong impression of the users' desires. The phrase 'want children' may be part of a sentence such as 'I do not want children', meaning that some of the 'want children' are false positives. That this is indeed the case can be seen by looking at the list of quadrigrams, where 'do not want children' (*buyao haizi* and *buyao xiaohai*) show up in 18th and 19th place. It turns out that 27 per cent of the references to 'want children' are extracted from phrases meaning 'do not want children'. It is important to be aware of such confounders when working with n-grams.

The frequency tables for 'want children' (Table 8) versus 'do not want children' (Table 9) further underline the gender difference on

the topic. While 'want children' appears to be equally present in male and female profiles, half of those mentions in the female profiles come from 'do not want children', while this is the case for less than one out of six mentions in male profiles.

Table 8: Frequency of 'want children'.

Frequency table: 'want children' (*yao haizi* 要孩子 and *yao xiaohai* 要小孩, combined)			
	Male users	Female users	All users
'Self' field	7,726 (17.7 times/100 profiles)	3,614 (19.1 times/100 profiles)	11,340 (18.1 times/100 profiles)
'Partner' field	5,411 (12.4 times/100 profiles)	2,465 (13.0 times/100 profiles)	7,876 (12.6 times/100 profiles)
Total	**13,137 (30.0 times/100 profiles)**	**6,079 (32.1 times/100 profiles)**	**19,216 (30.6 times/100 profiles)**

Table 9: Frequency of 'do not want children'.

Frequency table: 'do not want children' (*buyao haizi* 不要孩子 and *buyao xiaohai* 不要小孩, combined)			
	Male users	Female users	All users
'Self' field	1,315 (3.0 times/100 profiles)	1,708 (9.0 times/100 profiles)	3,023 (4.8 times/100 profiles)
'Partner' field	774 (1.8 times/100 profiles)	1,373 (7.2 times/100 profiles)	2,147 (3.4 times/100 profiles)
Total	**2,089 (4.8 times/100 profiles)**	**3,081 (16.2 times/100 profiles)**	**5,170 (8.2 times/100 profiles)**

In addition to deciding whether to have children, the topic of whether to cohabit is another crucial decision. Some *xinghun* seekers express a wish to remain largely autonomous after marriage and not be too involved in each other's lives. Mentioned 3,963 times in total, the phrase 'not interfere' (*bu ganshe*) is the 13th most frequent trigram. One male user writes: 'I hope we won't live together after marrying, except for in necessary circumstances'. A female user

writes: 'We won't meet a lot with [each other's] friends and family after marriage, only for the big holidays or if they have an issue or similar'.

For some users, *xinghun* appears to be mostly about performing a family role to the outside world. Some even explicitly use the word *xi*, meaning 'play' or 'show', to describe their marriage. A 31-year-old male user writes: 'For most of the time, we live with our own [boy/girl]friends. When necessary, we can help each other put on a show!' A 27-year-old female user, who mentions the phrase 'not interfere' three times in her profile, states: 'We shouldn't interfere in each other's lives. We shouldn't live together, because we [my girlfriend and I] are living together. […] Sometimes we will need to cooperate to put on a play'.

Table 10: Frequency of 'not interfere'.

Frequency table: 'not interfere' (*bu ganshe* 不干涉)			
	Male users	Female users	All users
'Self' field	1,190 (2.7 times/100 profiles)	767 (4.0 times/100 profiles)	1,957 (3.1 times/100 profiles)
'Partner' field	1,148 (2.6 times/100 profiles)	858 (4.5 times/100 profiles)	2,006 (3.2 times/100 profiles)
Total	**2,338 (5.3 times/100 profiles)**	**1,625 (8.6 times/100 profiles)**	**3,963 (6.3 times/100 profiles)**

The issue of 'non-interference' is brought up more often by female users than male users (Table 10). This may be related to the fact that a greater proportion of the female users on the site report to be in a same-sex relationship (Figure 5). As Engebretsen (2009: 12) points out, not being too involved in each other's lives—and especially, not having one's parents be too involved in one's life and marriage—can be an important factor in being able to have a same-sex partner and make room for one's homosexual identity. Indeed, one of the potential risks of *xinghun* might be that the spouses end up getting so involved with each other and their familial responsibilities that their romantic partners are put aside (Cho 2009; Wang 2014). Therefore, some may feel the need to establish clear boundaries before marriage.

Figure 5: Users' same-sex relationship status.

However, while some users talk about living apart and not interfering in each other's lives, another group expresses a wish to be emotionally involved with their partner. The word 'friend' (*pengyou*) is the 11th most frequent bigram, appearing 18,190 times, or 29 times per 100 profiles. A 28-year-old male user writes: 'I wish to find a *xinghun* partner that I can talk to, and where we can cooperate and understand each other—we can interact like friends'. The desire to engage in *xinghun* as a form of friendship can also be appreciated by noting that phrases such as 'good friend' (*hao pengyou*), 'like friends' (*xiang pengyou*) and 'be friends' (*zuo pengyou*) all appear quite frequently (2,190, 1,690 and 798 times, respectively).

Describing the partner as family is also a common trend. Two phrases meaning 'like family' (*xiang qinren* and *qinren yiyang*) appear 2,162 and 2,417 times, respectively (although 1,690 of them are overlapping occurrences, *xiang qinren yiyang*). A 32-year-old female user writes: 'Divorce shouldn't be the goal of *xinghun*. I hope we can interact like close family in the long term'. A 26-year-old male user gives an even more vivid description of the kind of long-term family relationship he hopes to have with his wife: 'I hope to find a person I can go on this journey with together, we can be friends, or even better, we can be family, close relatives. In many years, we may have lost our looks, and no one loves us anymore, but we can still go hand in hand'.

These statements show that some of those seeking a *xinghun* partner do not seek a superficial, performative agreement devoid of feelings or mutuality, but wish to establish emotional connections and commitments. This underlines Cho's (2009) perspective that *xinghun* is not a mere formality but creates a new social reality for the people involved.

Lastly, the concern for parents and a wish to live up to their expectations is a recurring theme in the profiles. 'Parents' (*fumu*) is the fourth most common bigram, while the top 25 quadrigrams include 'both parties' parents' (*shuangfang fumu*) and 'filial piety towards parents' (*xiaoshun fumu*). A 31-year-old male user writes: 'The only indispensable thing for me is my parents […] We can bring each other along to our families and put on a good show to make our parents happy'.

Filial piety is brought up equally frequently by male and female users (Table 11). The importance attributed to a filial partner is underscored by the fact that the phrase is much more common in the 'partner' field than the 'self' field. Given that China has customarily been a patrilocal society, it might have been expected for male users to emphasise filial piety in their future partner more than female users (Liu and Tan 2020). That this is not the case may testify to a development where married daughters maintain close ties with their biological parents after marriage.

Table 11: Frequency of 'filial piety towards parents'.

Frequency table: 'filial piety towards parents' (*xiaoshun fumu* 孝顺父母)			
	Male users	Female users	All users
'Self' field	807 (1.8 times/100 profiles)	357 (1.9 times/100 profiles)	1,164 (1.9 times/100 profiles)
'Partner' field	1,963 (4.5 times/100 profiles)	850 (4.5 times/100 profiles)	2,813 (4.5 times/100 profiles)
Total	**2,770 (6.3 times/100 profiles)**	**1,207 (6.4 times/100 profiles)**	**3,977 (6.3 times/100 profiles)**

Several users also verbalise the obligation to satisfy their parents as a form of 'pressure' (*yali*), which is mentioned no less than 7,075

times, though pressure is not only mentioned in relation to parents. A 31-year-old male user writes: 'I hope I can become friends with you and we can take the responsibility together, manage the pressure and justify ourselves to our parents and society', while a 35-year-old admits: 'Actually, *xinghun* is really my only option, we all do this for our parents'. Others, however, speak openly with their parents about the reality of marriage prospects. For example, a 30-year-old female user writes: 'I connect and relate to my parents without pressure'. As multiple scholars have noted, some gay people have an 'unspoken agreement' with their parents where everybody knows what is really going on—but as long as nobody says it out loud, they maintain the illusion that nothing is out of the ordinary (Engebretsen 2009: 4; Wang 2019: 30).

Conclusion

In this paper, we have examined which factors and motivations play into why young gay people in China choose to engage in *xinghun* with other gay people. Through a computational analysis of 62,710 dating profiles from the *xinghun* website Chinagayles.com, we have identified a wide range of themes and concerns that recur in gay/lesbian partner seeking.

The numerous mentions of parents in the profiles show that concern for their family is undoubtedly a major motivation for many *xinghun* seekers. Some users describe their prospective marriage as a performance, or directly state parental or societal pressure as an important motivation to engage in *xinghun*. A wish to conform to norms and expectations is further reflected in the widespread preference for traditional gender expressions and emphasis on finding a partner who is stable and not mingling with the 'wrong' crowd. For some gay people, *xinghun* offers a way to achieve stability and 'normality' and live up to the expectations set by their parents and society as a whole.

At the same time, however, our analysis shows that family pressure is not the only reason for Chinese gay people to engage in *xinghun*. The emphasis on financial independence seen in many profiles, as well as the frequent mentions of non-interference and keeping a distance between the spouses, show that many see *xinghun* as an opportunity for a more autonomous and independent life. Further, the recurrent emphasis on a partner with a kind and happy

personality and descriptions of one's future partner as a friend or family show that many do seek emotional support and friendship from a *xinghun* marriage and do not only enter into it as a façade.

Finally, the analysis shows that there are significant gender differences in the approach to *xinghun*. Male users generally want children, whereas many female users state an explicit wish to not have them. Instead, female users are more likely to have financial requirements for their potential husband and to seek a marriage of physical and emotional distance.

Overall, motivations for *xinghun* are many and varied and showcase a balancing act between complying with family norms and expectations while simultaneously being able to retain one's homosexual identity and desires. While our findings show that both pressure and personal desire play a role in the pursuit of *xinghun*, it is important to stress that these two dimensions cannot be neatly separated. Our desires are constantly reshaped by our social experiences, just as social norms change due to material, ideational and technological circumstances.

Another noteworthy conclusion of this study is that the simple procedure of working with n-grams can provide profound insights and perspectives. N-grams provide a straightforward and easily understandable method for revealing the most common words and phrases in Chinese text without relying on algorithmic decisions on what to count as a word and without searching for specific themes and expressions. Interpreting the significance of the n-gram frequencies still requires analytical skill and deep contextual knowledge, but it is noticeable that we can get quite a strong sense of what is contained in the data merely by investigating term frequencies. While manually reading through ten, fifty or a hundred profiles would enable us to notice many of the same topics, the aggregation of data represented by the n-gram frequency tables validates how widespread such issues are.

Since this study is based on profile texts from a dating website, it has certain limitations. What a person writes in a profile constitutes only a limited part of their complex considerations about *xinghun* and may be affected by dominant trends and discourses on the website. Since anyone can create a profile on the site, there is no way to verify the users or the information they provide. Further, the population surveyed has a significant over-representation of university educated people living in cities, and the findings of this study are thus representative only of users on Chinagayles.com and not of Chinese gay people in general. Keeping those limitations in mind, this study broadens the perspective on *xinghun* by contributing an extensive quantitative dimension and

demonstrates the usefulness of simple digital methods like analysing n-gram frequencies when working with large corpora of text.

ANNA DAVIDSEN BUHL is an MS student in IT & Cognition at the University of Copenhagen with a BA in China Studies. E-mail: zkn227@ alumni.ku.dk

BO ÆRENLUND SØRENSEN is a Tenure-Track Assistant Professor at the University of Copenhagen, specialising in research on Sinophone culture and media. E-mail: bo.aerenlund@hum.ku.dk

ENDNOTES

1   Chinagayles.com. 'Guanyu women' (About us). https://www. chinagayles.com/about.php. Accessed December 2024.
2   The law, abolished in 1997, was against 'hooliganism' and did not explicitly mention homosexuality, but had been used to prosecute gay people.
3   For more on n-grams, see Jurafsky and Martin 2024, Chapter 3; Silge and Robinson 2017, Chapter 4.
4   See https://github.com/fxsjy/jieba. Accessed December 2024.
5   Although technically a verb, 'to feminise', nüxinghua is used in the profiles to simply mean 'feminine'.

REFERENCES

Altehenger, Jennifer E. 2018. *Legal Lessons: Popularizing Laws in the People's Republic of China, 1949-1989*. Cambridge, MA: Harvard University Asia Center Publications Program. https://doi.org/10. 1080/09668136.2020.1725302.

Bao, Hongwei 2018. *Queer Comrades: Gay Identity and Tongzhi Activism in Postsocialist China*. Copenhagen: NIAS Press.

——— 2020. *Queer China: Lesbian and Gay Literature and Visual Culture Under Postsocialism*. New York: Routledge. https://doi. org/10.4324/9781003027898.

Barbalet, Jack 2015. 'Guanxi, Tie Strength, and Network Attributes'. *American Behavioral Scientist* 59 (8): 1038-1050. https://doi. org/10.1177/0002764215580613.

Bird, Steven, Ewan Klein and Edward Loper 2019. *Natural Language Processing with Python: Analyzing Text with the Natural Language*

Toolkit. https://www.nltk.org/book/. Accessed December 2024.

Chang, Leslie T. 2009. *Factory Girls: From Village to City in a Changing China.* New York: Random House.

Chen, Yaya 2009. 'Nü tongxinglianzhe de hunyin he jiating gei chuantong hunyin zhidu dailai de tiaozhan' (The Challenges Lesbian Women's Marriage and Families Bring to the Traditional Marriage System). *Shehui zazhi* (Society) 29 (4): 107-129.

Cho, John (Song Pae) 2009. 'The Wedding Banquet Revisited: "Contract Marriages" Between Korean Gays and Lesbians'. *Anthropological Quarterly* 82 (2): 401-422. https://doi.org10.1353/anq.0.0069.

Choi, Susanne YP and Ming Luo 2016. 'Performative Family: Homosexuality, Marriage and Intergenerational Dynamics in China'. *The British Journal of Sociology* 67 (2): 260-280. https://doi.org/10.1111/1468-4446.12196.

Engebretsen, Elisabeth Lund 2009. 'Intimate Practices, Conjugal Ideals: Affective Ties and Relationship Strategies Among *Lala* (Lesbian) Women in Contemporary Beijing'. *Sexuality Research & Social Policy* 6 (3): 3-14. https://doi.org/10.1525/srsp.2009.6.3.3.

Engebretsen, Elisabeth Lund 2017. 'Under Pressure: Lesbian-Gay Contract Marriages and Their Patriarchal Bargains'. In Gonçalo Santos and Stevan Harrell (eds.) *Transforming Patriarchy: Chinese Families in the Twenty-First Century.* Seattle: University of Washington Press, pp. 163-181. https://doi.org/10.1515/9780295998985-011.

Firth, John Rupert 1957. 'A Synopsis of Linguistic Theory 1930-1955'. In John R. Firth (ed.) *Studies in Linguistic Analysis.* Oxford: Blackwell, pp. 1-32.

Grimmer, Justin, Margaret E. Roberts and Brandon M. Stewart 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences.* Princeton: Princeton University Press.

Gui, Tianhan and Ke Meng 2023. 'Doing Gender, Performing Marriage: A Study of China's *Xinghun* Marriage Ads'. *Women's Studies International Forum* 98: 1-10. https://doi.org/10.1016/j.wsif.2023.102718.

Hart, Roderick P., Sharon E. Jarvis, William P. Jennings and Deborah Smith-Howell 2005. *Political Keywords: Using Language That Uses Us.* New York: Oxford University Press.

Huang, Shuzhen 2018. 'Beyond the Sex–Love–Marriage Alignment: *Xinghun* Among Queer People in Mainland China'. In Michael W. Yarbrough, Angela Jones and Joseph N. DeFilippis (eds.) *Queer Families and Relationships After Marriage Equality.* London: Routledge, pp. 136-149.

Jankowiak, William R. and Robert L. Moore 2016. *Family Life in China*. Cambridge: Polity Press. https://doi.org/10.1111/1467-9655.12897.

Jurafsky, Daniel and James H. Martin 2024. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd edition. https://web.stanford.edu/~jurafsky/slp3. Accessed December 2024.

Kam, Lucetta Yip Lo 2013. *Shanghai Lalas: Female Tongzhi Communities and Politics in Urban China*. Hong Kong: Hong Kong University Press. https://doi.org/10.5790/hongkong/9789888139453.001.0001.

Liu, Min 2013. 'Two Gay Men Seeking Two Lesbians: An Analysis of Xinghun (Formality Marriage) Ads on China's Tianya.cn'. *Sexuality & Culture* 17 (3): 494-511. https://doi.org/10.1007/s12119-012-9164-z.

Liu, Tingting and Chris K. K. Tan 2020. 'On the Transactionalisation of Conjugal Bonds: A Feminist Materialist Analysis of Chinese Xinghun Marriages'. *Anthropological Forum* 30 (4): 443-463. https://doi.org/10.1080/00664677.2020.1855108.

Mendos, Lucas Ramón et al. 2020. 'State-Sponsored Homophobia 2020: Global Legislation Overview Update'. ILGA World.

National Bureau of Statistics of China 2022. 'Population at Year-end by Region'. In *China Statistical Yearbook 2022*. https://www.stats.gov.cn/sj/ndsj/2022/indexeh.htm. Accessed September 2024.

NPC Standing Committee 2018. *Law of the People's Republic of China on Protection of the Rights and Interests of the Elderly* (2018 Amendment).

Osburg, John 2024. 'Chinese Consumer Culture: Luxury, Nationalism, and Class Formation.' In Mayfair Yang (ed.) *Anthropology of Ascendant China: Histories, Attainments, and Tribulations*. London: Routledge, pp. 89-107. https://doi.org/10.4324/9781003358947.

Ren, Xiaoyi and Tianhan Gui 2022. 'Where the Rainbow Rises: The Strategic Adaptations of China's LGBT NGOs to Restricted Civic Space'. *Journal of Contemporary China* 33 (145): 101-119. https://doi.org/10.1080/10670564.2022.2131378.

Sangren, Kristin 2021. 'Adjudicating Filial Piety: Law, Legitimacy, and Moral Authority in the PRC', AAS Annual Conference, March 21-26, Online.

Silge, Julia and David Robinson 2017. *Text Mining with R: A Tidy Approach*. Sebastopol: O'Reilly Media.

Song, Changhui, Joyee Shairee Chatterjee, Donna L. Doane and Philippe Doneys 2023. 'Men in Mixed-Orientation Marriages in

Contemporary China: Unpacking the Role of Heteronormativity and Patriarchy'. *Journal of Family Issues* 44 (5): 1197-1214. https://doi.org/10.1177/0192513X211055790.

Underwood, Ted 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226612973.001.0001.

United Nations Development Programme 2016. *Being LGBTI in China: A National Survey on Social Attitudes Towards Sexual Orientation, Gender Identity and Gender Expression*.

Wang, Stephanie Yingyi 2019. 'When Tongzhi Marry: Experiments of Cooperative Marriage between Lalas and Gay Men in Urban China'. *Feminist Studies* 45 (1): 13-35. https://doi.org/10.1353/fem.2019.0007.

Wang, Yingyi 2014. 'Hezuo hunyin chushen: Nannü tongzhi de hunyin dongji yanjiu' (Exploring Cooperative Marriage: A Study of Motivations for Marriages Between Gay Men and Lesbian Women). *Zhongguo nianqing yanjiu* (China Youth Study) 11: 14-18.

Yan, Yunxiang 2018. 'Neo-Familism and the State in Contemporary China'. *Urban Anthropology and Studies of Cultural Systems and World Economic Development* 47 (3): 181-224.

Yu, Jia and Xie, Yu 2015. 'Changes in the Determinants of Marriage Entry in Post-Reform Urban China'. *Demography* 52 (6): 1869-1892. https://doi.org/10.1007/s13524-015-0432-z.

Zhong, Juan and Jeffrey J. Arnett 2018. '"I Felt Like There Would Be a Vast Sky Out There": The Leaving Home Decisions of Chinese Migrant Women Workers'. *International Perspectives in Psychology* 7 (3): 121-136. https://doi.org/10.1037/ipp0000088.