



# تحلیل شبکه‌ای سایت‌های میزبانی کد مطالعه موردی سایت گیت‌لب

مهندس هادی صفری

دانش‌آموخته دانشگاه تهران

انتشارات میعاد اندیشه

۱۳۹۹



انتشارات میعاد اندیشه

سرشناسه	: صفری، هادی، ۱۳۷۶ -
عنوان و نام پدیدآور	: تحلیل شبکه‌ای سایت‌های میزبانی کد: مطالعه موردی سایت گیت‌لب / هادی صفری.
مشخصات نشر	: تهران: میعاد اندیشه، ۱۳۹۹.
مشخصات ظاهری	: ج، ۱۰۲ ص.: مصور، جدول، نمودار.
شابک	: ۹۷۸-۶۲۲-۲۳۱۳-۱۶-۶
وضعیت فهرست نویسی	: فیپا
یادداشت	: کتابنامه: ص. [۸۳] - ۱۰۲.
موضوع	: سایت اینترنتی گیت‌لب
موضوع	: GitLab
موضوع	: وبگاه‌ها
موضوع	: Web sites
رده بندی کنگره	: TK۵۱۰۵/۸۸۲
رده بندی دیویی	: ۰۰۴/۶۷۸
شماره کتابشناسی ملی	: ۶۲۳۹۰۷۸

تحلیل شبکه‌ای سایت‌های میزبانی کد: مطالعه موردی سایت گیت‌لب  
مؤلف: هادی صفری  
صفحه‌آرا: هادی صفری  
طراح جلد: علی نظام‌دوست  
ناشر: میعاد اندیشه  
تاریخ و نوبت چاپ: ۱۳۹۹ - اول  
شمارگان: ۵۰  
قیمت نسخه چاپی: ۲۵,۰۰۰ تومان  
قیمت نسخه الکترونیکی: ۸,۰۰۰ تومان  
شابک نسخه چاپی: ۹۷۸-۶۲۲-۲۳۱-۳۱۶-۶  
شابک نسخه الکترونیکی: ۹۷۸-۶۲۲-۲۳۱-۳۶۸-۵  
چاپ: پیشگام  
پخش نسخه الکترونیکی: <https://taaghche.com> (طاقچه)

برای خرید اینترنتی به سایت انتشارات میعاد اندیشه (<https://www.chapemiaad.com>) مراجعه کنید.  
تمامی حقوق برای نویسنده محفوظ است.

# پیش‌گفتار

امروزه برنامه‌نویسان تسهیلات میزبانی کد را با گیت و سایت گیت‌هاب می‌شناسند. گیت زیرساخت‌های لازم را برای برنامه‌نویسی چندنفره و نگهداری تاریخچه کدها فراهم می‌کند. اما با وجود محبوبیت گسترده گیت‌هاب، این سایت تنها ارائه‌دهنده خدمات میزبانی کد نیست. سایت گیت‌لب در سال‌های اخیر محبوبیت رو به رشدی داشته است و توانسته است مقام دوم را از نظر تعداد مخازن در بین سایت‌های رقیب کسب کند. سرعت بالای گسترش گیت‌لب و فقدان تحقیقات کافی درباره مخازن و کاربران آن نگارنده را به بررسی سایت مذکور ترغیب کرد. اثر حاضر که برگرفته از پایان‌نامه دوره کارشناسی نگارنده در دانشگاه تهران برای کسب درجه علمی مهندسی نرم‌افزار است، نتایج این پژوهش را در بر می‌گیرد.

در ابتدا با جمع‌آوری فراداده پروژه‌های موجود در گیت‌لب با یک خزنده به زبان پایتون، یک پایگاه‌داده MySQL از اطلاعات پروژه‌ها و کاربران موجود در این سایت میزبانی کد تشکیل شد. در مرحله بعد، اطلاعات جمع‌آوری شده از جنبه‌های مختلف، با تأکید بر روش‌های گرافی رایج در تحلیل شبکه‌های اجتماعی و نیز تحلیل‌های آماری، مورد تحلیل و بررسی قرار گرفتند. برای بیشتر تحلیل‌ها از زبان پایتون و کتابخانه نتورک‌اکس و برای مصورسازی داده‌ها از گفنی استفاده شد؛ هر چند ابزارهای دیگری نیز بسته به نیاز مورد استفاده قرار گرفتند.

روش‌های کلاسیکی مانند بررسی ویژگی‌های ساختاری گراف‌ها، بررسی شاخص‌های مرکزیت، تشکیل گراف نمایش وزن‌دار ارتباطات کاربران و نیز پروژه‌ها و تحلیل آن و بررسی همبستگی ویژگی‌های مختلف پروژه‌ها منجر به یافتن اطلاعات جالب توجهی هم درباره کلیت جامعه کاربران و پروژه‌های سایت گیت‌لب و هم درباره مهم‌ترین کاربران و پروژه‌های آن سایت شده است. بسیاری از ویژگی‌های ساختاری شبکه‌های مورد بررسی سایت گیت‌لب با نتایج حاصل از تحقیقات موجود درباره سایت‌های دیگر میزبانی کد مشابهت داشت. همچنین، مهم‌ترین و تأثیرگذارترین پروژه‌ها و کاربران سایت به نوعی با خود تیم گیت‌لب و پروژه‌های زیرمجموعه آن مرتبط بوده‌اند؛ به نظر می‌رسد هنوز برنامه‌نویسان زیادی از سایت مذکور برای توسعه پروژه‌های گروهی و بزرگ استفاده نمی‌کنند. در طول این پژوهش اشکالاتی نیز در سایت مذکور یافت شد که به اطلاع پشتیبانی سایت رسید.

نتایج تحقیق حاضر، علاوه بر آن که می‌تواند به پژوهشگران دید جامعی نسبت به سایت گیت‌لب و روند تغییرات الگوهای رفتاری توسعه‌دهندگان مستقل بدهد و مقدمه‌ای بر پژوهش‌های عمیق‌تر درباره این سایت باشد، به گردانندگان و سیاست‌گذاران سایت مذکور نیز در شناخت نیازمندی‌های کاربران و برنامه‌ریزی‌های فنی و تجاری برای افزایش رضایت مشتریان، ارائه امکانات مؤثرتر و پیشی گرفتن از رقبا یاری خواهد رساند.

در پایان، لازم می‌دانم از زحمات استاد گرامی‌ام، دکتر بهنام بهرک، که در طول این پژوهش مشوق و راهنمای من بودند، پدرم، دکتر جهانگیر صفری، که سال‌ها پیش مرا با تحقیق و پژوهش آشنا کردند، علی‌نظام‌دوست که زحمت طراحی جلد را متقبل شدند، مادرم، سارا جهانگیری و مهندس سید محمدرضا طیرانیان حسینی که مرا با خواندن نمونه اولیه کتاب یاری کردند نهایت تشکر را به عمل آورم.

هادی صفری

[hadi.safari@ut.ac.ir](mailto:hadi.safari@ut.ac.ir)

دانشگاه تهران

# فهرست مطالب

پ	پیش‌گفتار
۱	۱ مقدمه
۵	۲ اصطلاحات و تعاریف
۵	۱.۲ اصطلاحات مربوط به گیت‌لب
۶	۲.۲ اصطلاحات مربوط به گراف
۷	۱.۲.۲ شاخص‌های مرکزیت
۸	۳.۲ اصطلاحات آماری و سایر اصطلاحات
۹	۳ مروری بر مطالعات پیشین
۱۰	۱.۳ گیت‌لب
۱۱	۲.۳ گیت‌هاب
۴۵	۳.۳ سورس‌فوریج
۴۹	۴.۳ گوگل‌کد

۴۹	سایر تحقیقات	۵.۳
۵۱	روش تحقیق	۴
۵۱	جمع‌آوری داده‌ها	۱.۴
۵۴	تحلیل داده‌ها	۲.۴
۵۴	تحلیل انشقاق‌ها	۱.۲.۴
۵۴	تحلیل گراف دوبخشی افراد و پروژه‌ها	۲.۲.۴
۵۵	تحلیل ویژگی‌های پروژه‌ها	۳.۲.۴
۵۷	نتایج	۵
۵۷	تحلیل انشقاق‌ها	۱.۵
۶۲	تحلیل گراف دوبخشی افراد و پروژه‌ها	۲.۵
۶۲	گراف مشارکت‌ها	۱.۲.۵
۶۴	گراف کاربران	۱.۱.۲.۵
۶۴	گراف پروژه‌ها	۲.۱.۲.۵
۷۱	گراف عضویت	۲.۲.۵
۷۳	تحلیل ویژگی‌های پروژه‌ها	۳.۵
۷۳	همبستگی ویژگی‌های پروژه‌ها	۱.۳.۵
۷۳	توزیع ویژگی‌های پروژه‌ها	۲.۳.۵
۷۹	فرآورده‌های فرعی پژوهش	۴.۵
۸۱	بحث و نتیجه‌گیری	۶
۸۳	منابع	



## مقدمه

توسعه گروهی و از راه دور نرم‌افزارها روز به روز به موضوعی پرطرفدارتر میان برنامه‌نویسان، خصوصاً برنامه‌نویسان متن‌باز [۱۰۷، ص. ۸۰] تبدیل می‌شود. همچنین، نگهداری سابقه توسعه نرم‌افزار برای بررسی روند پیشرفت کار و نیز تسهیل عیب‌یابی نرم‌افزارها اهمیت روزافزون پیدا کرده است. از سال‌ها پیش، نرم‌افزارهای کنترل نسخه توزیع شده<sup>۱</sup> این دو مسأله را ترکیب کرده‌اند تا زیرساخت مناسبی برای پاسخ‌گویی به هر دو نیاز فراهم کنند. امروزه استفاده از این سیستم‌ها مختص برنامه‌نویسان نیست و گروه‌های دیگری مانند گرافیک‌ها نیز به جمع کاربران این سیستم‌ها پیوسته‌اند. از این سیستم‌ها حتی می‌توان برای توسعه اسناد متنی نیز استفاده کرد [۱۳۶]. میزبانی سایت‌های ایستا در برخی از سایت‌های میزبان مخازن کد امکان وبلاگ‌نویسی را نیز با استفاده از سیستم‌های کنترل نسخه فراهم کرده است [۲۵۷].

بعد از آن که لینوس توروالدز<sup>۲</sup> در سال ۲۰۰۵ گیت<sup>۳</sup> را به عنوان ابزار جدیدی برای توسعه کد لینوکس<sup>۴</sup> معرفی کرد، این سیستم کنترل نسخه به سرعت فراگیر شد و امروزه بیشترین استفاده

---

<sup>1</sup>Distributed Version Control Systems    <sup>2</sup>Linus Benedict Torvalds    <sup>3</sup>git    <sup>4</sup>Linux

## جدول ۱.۱ محبوب‌ترین سایت‌های ارائه‌دهنده تسهیلات میزبانی کد به ترتیب تعداد مخازن

[Popularity, ۲۴۵]		
نام	تعداد کاربران	تعداد مخازن
گیت‌هاب	۲۴,۰۰۰,۰۰۰	۶۹,۰۰۰,۰۰۰
گیت‌لب	۱۰۰,۰۰۰	۹۵۴۶,۰۰۰ <sup>۹</sup>
اسمبلا	؟	۵۲۶,۵۸۱
سورس‌فورج	۳,۷۰۰,۰۰۰	۵۰۰,۰۰۰
لانچ‌پد	۳,۹۶۵,۲۸۸	۴۰,۸۸۱
بیت‌باکت	۵,۰۰۰,۰۰۰	؟

را در میان سیستم‌های مشابه دارد [۲۶۲، ص ۴۵۹]. [برای آشنایی بیشتر با گیت ر.ک. [۵۴، ۳۰، ۱۷۹، ۱۸۰].]

گیت‌لب<sup>۱</sup> [۷۸] یک سایت ارائه‌دهنده تسهیلات توسعه نرم‌افزار است که امکان میزبانی کد را تحت مخازن گیت فراهم می‌کند. گیت‌لب را سیتسه سیراندی<sup>۲</sup> و دیمیتری زاپروژتس<sup>۳</sup> در سال ۲۰۰۷ راه‌اندازی کردند و آن را از سانفرانسیسکو آمریکا مدیریت می‌کنند [۲۴۶]. پرکاربردترین زبان مورد استفاده برای توسعه سایت روبی<sup>۴</sup> بوده است. این سایت همانند سایت‌های رقیب، علاوه بر میزبانی مخازن کد، امکانات دیگری را مانند تسهیلات مرور کد<sup>۵</sup>، CI/CD<sup>۶</sup>، مخزن تصاویر داکر<sup>۷</sup> و میزبانی ویکی نیز فراهم می‌کند.

گیت‌لب در سال‌های اخیر بسیار محبوب شده است و بر خلاف شش سال پیش [۲۶۲، ص ۴۵۹]، امروزه از نظر تعداد مخازن پس از گیت‌هاب<sup>۸</sup> در جایگاه دوم قرار دارد [جدول ۱.۱]. پس از آن که در سال ۲۰۱۸ مایکروسافت گیت‌هاب را خرید، موج مهاجرت کاربران از گیت‌هاب به گیت‌لب خبرساز شد [۱۷۳]. افزایش سریع تعداد مخازن سایت در طول مدت جمع‌آوری داده‌های لازم برای این پژوهش نیز قابل مشاهده بود.

تحقیقات بسیاری بر روی جنبه‌های مختلف سایت‌های میزبانی کد انجام شده است. برخی از این تحقیقات به ویژگی‌های کدهای موجود در مخازن پرداخته‌اند. برخی دیگر از نظر متدولوژی‌های توسعه نرم‌افزار به بررسی سیستم‌های کنترل نسخه و سایر امکانات این سایت‌ها پرداخته‌اند. پژوهش‌های دیگری سایت‌های میزبانی کد را به عنوان یک شبکه اجتماعی مورد

<sup>۹</sup>اطلاعات ۶۶۷,۶۸۶ مخزن در فاز جمع‌آوری اطلاعات پژوهش حاضر جمع‌آوری شده است. به قسمت پروژه‌ها مراجعه کنید.

<sup>۱</sup>GitLab    <sup>۲</sup>Sytse Sijbrandij    <sup>۳</sup>Dmitriy Zaporozhets    <sup>۴</sup>Ruby    <sup>۵</sup>code reveiw  
<sup>۶</sup>Continuous Integration/Deployment    <sup>۷</sup>Docker    <sup>۸</sup>GitHub



بررسی قرار داده‌اند. پژوهش‌های قدیمی‌تر بر روی سایت سورس‌فورج<sup>۱</sup> و پژوهش‌های جدیدتر بر روی سایت گیت‌هاب متمرکز بوده‌اند. تحقیقاتی نیز دربارهٔ سایت گیت‌لب یافت می‌شود، اما بیشتر آن‌ها به تأثیر استفاده در گیت‌لب در آموزش کامپیوتر پرداخته‌اند.

در پژوهش حاضر، فصل **مقدمه** به دلایل اهمیت سایت گیت‌لب و بررسی جنبه‌های اجتماعی این سایت اختصاص یافته است. در فصل **اصطلاحات و تعاریف** به اختصار برخی اصطلاحات و تعاریف مورد استفاده شرح داده شده‌اند. فصل **مروری بر مطالعات پیشین** به پیشینهٔ تحقیق و بررسی مقالات و پژوهش‌های مشابه اختصاص یافته است. در فصل **روش تحقیق** روش جمع‌آوری داده‌ها و تفسیر آن‌ها مورد بررسی قرار گرفته است. در فصل **نتایج** نتایج تحلیل داده‌ها به تفصیل شرح داده شده است. فصل **بحث و نتیجه‌گیری** نیز به جمع‌بندی تحقیقات و نتیجه‌گیری از آن‌ها اختصاص یافته است.

---

<sup>1</sup>SourceForge



# ۲

## اصطلاحات و تعاریف

در این فصل برخی اصطلاحات و تعاریف‌های استفاده‌شده در پروژه شرح داده شده‌اند.

### ۱.۲ اصطلاحات مربوط به گیت‌لب

ستاره<sup>۱</sup> کاربران می‌توانند به مخازن و موضوعات ستاره دهند تا مسائل آن‌ها را دنبال کنند و در خوراک<sup>۲</sup> خبر خود آن‌ها را مشاهده کنند.

انشقاق<sup>۳</sup> یک رونوشت از یک مخزن نرم‌افزاری در حوزه نام یک کاربر یا یک گروه دیگر است. در گیت‌هاب و گیت‌لب این کار معمولاً برای توسعه مخزن اصلی کاربرد دارد و سازنده محسوب می‌شود، اما در سورس‌فورج معمولاً برای توسعه یک نسخه متفاوت به کار می‌رفت و به معنی ظهور یک رقیب جدید بود.

مراقبت<sup>۴</sup> یک کاربر می‌تواند یک پروژه را مراقبت کند تا اعلانات مربوط به همه اتفاقات آن پروژه را دریافت کند.

---

<sup>0</sup>star   <sup>2</sup>feed   <sup>2</sup>fork   <sup>3</sup>watch

**اشکال<sup>۱</sup>** کاربران می‌توانند اشکالات پروژه‌ها را در بخش مخصوصی گزارش دهند. این گزارش‌ها ممکن است به توسعه‌دهندگانی نسبت داده شوند تا به آن‌ها پاسخ گویند یا اشکال را برطرف کنند.

**درخواست ترکیب<sup>۲</sup>** در مدل توسعه نرم‌افزار گیت‌هاب و گیت‌لب، توسعه‌دهندگان جزء یا افراد متفرقه بعد از توسعه کد بر روی انشقاق خودشان از پروژه، یک درخواست ترکیب در پروژه اصلی ایجاد می‌کنند. مالکان پروژه یا توسعه‌دهندگان اصلی پروژه با مشاهده این درخواست‌های ترکیب آن‌ها را تأیید یا رد می‌کنند یا از درخواست‌دهنده می‌خواهند تغییراتی ایجاد کند. در صورت تأیید درخواست، تغییرات در مخزن اصلی نیز اعمال می‌شوند.

**ارتکاب<sup>۳</sup>** به معنی ثبت مجموعه‌ای از تغییرات در نسخه محلی مخزن است. هر ارتکاب یک پیام دارد که تغییرات آن را توصیف می‌کند.

## ۲.۲ اصطلاحات مربوط به گراف

**همسان‌گزینی<sup>۴</sup>** تمایل گره‌های یک شبکه برای اتصال به گره‌های مشابه است.

**ضریب باشگاه اغنیا<sup>۵</sup>** شاخصی در گراف‌هاست که نشان می‌دهد گره‌های درجه بالا چقدر تمایل دارند به یکدیگر متصل شوند.

**شبکه مستقل از مقیاس<sup>۶</sup>** شبکه‌ای است که توزیع درجه آن از قانون توانی<sup>۷</sup> پیروی می‌کند. این نوع شبکه‌ها گره‌های مرکزی<sup>۸</sup> با درجه‌های بسیار بالاتر از میانگین دارند.

**شبکه‌های خودمحور<sup>۹</sup>** شبکه‌ای است که حول یک شخص شکل گرفته است.

---

<sup>۴</sup>issues   <sup>۱</sup>pull request   <sup>۲</sup>commit   <sup>۳</sup>assortativity   <sup>۴</sup>rich-club coefficient   <sup>۵</sup>scale-free networks  
<sup>۷</sup>power law   <sup>۸</sup>hub   <sup>۹</sup>ego-centric network

### ۱.۲.۲ شاخص‌های مرکزیت

«مرکزیت یک فرد در شبکه اجتماعی نشان‌دهنده پرستیژ و اقتدار آن فرد در شبکه است [۲].»  
مرکزیت شاخصی برای کمی‌سازی برتری نقش‌آفرینی افراد در شبکه است.  
شاخص‌های مختلفی برای بررسی مرکزیت گره‌های یک شبکه به کار می‌رود.

مرکزیت درجه<sup>۱</sup> این شاخص مرکزیت درجه گره‌ها را در نظر می‌گیرد. در گراف‌های جهت‌دار، رتبه خروجی بالا نشانگر نفوذ و تأثیر زیاد در شبکه و رتبه ورودی بالا نشانگر اعتبار یا شهرت است [۲].

مرکزیت نزدیکی<sup>۲</sup> این شاخص مرکزیت نشان می‌دهد یک موجودیت در شبکه چقدر سریع می‌تواند به موجودیت‌های بیشتری در شبکه دسترسی پیدا کند [۲].

$$C(x) = \frac{1}{\sum_y d(y, x)}$$

مرکزیت بینابینی<sup>۳</sup> این شاخص مرکزیت تأثیر افراد را بر روی جریان اطلاعات بین دیگران می‌سنجد [۲].

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

مرکزیت بردار ویژه<sup>۴</sup> این شاخص مرکزیت بر این اساس کار می‌کند که ارتباط با گره‌های با مرکزیت بالا، در مقایسه با ارتباط با گره‌های با مرکزیت پایین، قدرت بیشتری به دنبال دارد [۲].

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

<sup>۱</sup>degree centrality    <sup>۲</sup>closeness centrality    <sup>۳</sup>betweenness centrality    <sup>۴</sup>eigenvector centrality

مرکزیت کاتز<sup>۱</sup> این شاخص مرکزیت نسخه‌ای از مرکزیت بردار ویژه است که برخی مشکلات آن را، از جمله نشت مرکزیت از شبکه، با افزودن یک مقدار اولیه برطرف می‌کند.

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

### ۳.۲ اصطلاحات آماری و سایر اصطلاحات

توزیع دم‌سنگین<sup>۲</sup> به توزیع‌هایی گفته می‌شود که دم توزیع آن‌ها سنگین‌تر از دم توزیع نمایی<sup>۳</sup> است. در بسیاری از کاربردها توزیع دم راست مورد علاقه است با این حال ممکن است توزیع چپ یا هر دو طرف دم توزیع (راست و چپ) سنگین باشند.

قانون پارتو<sup>۴</sup> که با نام‌های قانون ۸۰-۲۰، قانون افراد اندک اساسی<sup>۵</sup> و اصل تنکی فاکتور<sup>۶</sup> نیز شناخته می‌شود بیان می‌کند که ۸۰ درصد رخدادهای ۲۰ درصد دلایل به وجود می‌آید.

شاخص کامیون<sup>۷</sup> یا شاخص اتوبوس<sup>۸</sup> حداقل تعداد افرادی را نشان می‌دهد که اگر با یک کامیون یا اتوبوس تصادف کنند (یا از پروژه خارج شوند) پروژه از کار می‌افتد.

<sup>۱</sup>Katz centrality    <sup>۲</sup>heavy-tailed distribution    <sup>۳</sup>exponential distribution    <sup>۴</sup>Pareto principle  
<sup>۵</sup>law of the vital few    <sup>۶</sup>principle of factor sparsity    <sup>۷</sup>truck factor    <sup>۸</sup>bus factor



**Network Analysis  
of  
Source Code Hosting Websites  
GitLab Case Study**

By  
**Hadi Safari**

**Miaad Andishe Publication**  
2020